(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/37222 A1

(51) International Patent Classification[7]: G06T 17/00, G06K 9/00

(21) International Application Number: PCT/GB00/04411

(22) International Filing Date:
20 November 2000 (20.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9927314.6 18 November 1999 (18.11.1999) GB

(71) Applicant *(for all designated States except US)*: ANTHROPICS TECHNOLOGY LIMITED [GB/GB]; Ealing Studios, Ealing Green, London W5 5EP (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: NEWMAN, Rhys, Andrew [GB/GB]; Anthropics Technology Limited, Ealing Studios, Ealing Green, London W5 5EP (GB). WILES, Charles, Stephen [GB/GB]; Anthropics Technology Limited, Ealing Studios, Ealing Green, London W5 5EP (GB). WILLIAMS, Mark, Jonathan [GB/GB]; Anthropics Technology Limited, Ealing Studios, Ealing Green, London W5 5EP (GB).

(74) Agents: BERESFORD, Keith, Denis, Lewis et al.; Beresford & Co, 2-5 Warwick Court, High Holborn, London WC1R 5DJ (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

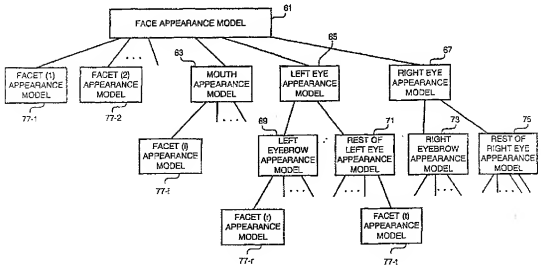
Published:

— *With international search report.*

— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IMAGE PROCESSING SYSTEM



(57) Abstract: A hierarchical parametric model is provided for modelling the appearance of objects, such as human faces. The hierarchical model can model both the shape and the texture of the object. The hierarchical model includes models for components of the object being modelled such that output parameters from one model are applied as input parameters to models which are lower in the hierarchy. This hierarchical model can be used, for example, for face tracking, for video compression, for 2D and 3D character

WO 01/37222 A1

IMAGE PROCESSING SYSTEM

The present invention relates to the parametric modelling of the appearance of objects. The resulting model can be used, for example, to track the object, such as a human face, in a video sequence.

The use of parametric models for image interpretation and synthesis has become increasing popular. Cootes et al have shown in their paper entitled "Active Shape Models - Their Training and Application", Computer Vision and Image Understanding, Volume 61, No. 1, January, pages 38-59, 1995, how such parametric models can be used to model the variability of the shape and texture of human faces. They have mainly used these models for face recognition and tracking within video sequences, although they have also demonstrated that their model can be used to model the variability of other deformable objects, such as MRI scans of knee joints. The use of these models provides a basis for a broad range of applications since they explain the appearance of a given image in terms of a compact set of model parameters which can be used for higher levels of interpretation of the image. For example, when analysing face images, they can be used to characterise the identity, pose or expression of a face.

Using such models for image interpretation requires, however, a method of fitting them to new image data.

This involves identifying the model parameters that generate an image which best fits (according to some measure) the new input image. Typically this problem is one of minimising the sum of squares of pixel errors between the generated image and the input image. In their paper entitled "Estimating Coloured 3D Face Models from Single Images: An Example-Based Approach" Vetter and Blanz have proposed a stochastic gradient descent optimisation technique to identify the optimum model parameters for the new image. Although this technique can give very accurate results finding the locally optimal solution, they generally get stuck in local minima since the error surface for the problem of fitting an appearance model to an image is particularly rough containing many local minima. Therefore, this minimisation technique often fails to converge on the global minimum. An additional drawback of this technique is that it is very slow requiring several minutes to achieve convergence.

A faster, more robust technique known as the active appearance model was proposed by Edwards et al in the paper entitled "Interpreting Face Images using Active Appearance Models", published in the Third International Conference on Automatic Face and Gesture Recognition 1998, pages 300-305, Japan, April 1998. This technique uses a prior training stage in which the relationship between model parameter displacements and the resulting

change in image error is learnt. Although the method is much faster than direct optimisation techniques, it also requires fairly accurate initial model parameters if the search is to converge. Additionally, this technique does not guarantee that the optimum parameters will be found.

The appearance model proposed by Cootes et al includes a single appearance model matrix which linearly relates a set of parameters to corresponding image data. Blanz et al segmented the face into a number of completely independent appearance models, each of which is used to render a separate region of the face. The results are then merged using a general interpretation technique.

The present invention aims to provide an alternative way of modelling the appearance of objects which will allow subsequent image interpretation through appropriate processing of parameters generated for the image.

According to one aspect, the present invention provides a hierarchical parametric model for modelling the shape of an object, the model comprising data defining a hierarchical set of functions in which a function in a top layer of the hierarchy is operable to generate a set of output parameters from a set of input parameters and in which one or more functions in a bottom layer of the hierarchy are operable to receive parameters output from one or more functions from a higher layer of the

hierarchy and to generate therefrom the relative positions of a plurality of predetermined points on the object. Such a hierarchical parametric model has the advantage that small changes in some parts of the object can still be modelled by the parameters, even though they are significantly smaller than variations in other less important parts of the object. This model can be used for face tracking, video compression, 2D and 3D character generation, face recognition for security purposes, image editing etc.

According to another aspect, the present invention provides an apparatus and method of determining a set of appearance parameters representative of the appearance of an object, the method comprising the steps of storing a hierarchical parametric model such as the one discussed above and at least one function which relates a change in input parameters to an error between actual appearance data for the object and appearance data determined from the set of input parameters and the parametric model; initially receiving a current set of input parameters for the object; determining appearance data for the object from the current set of input parameters and the stored parametric model; determining the error between the actual appearance data of the object and the appearance data determined from the current set of input parameters; determining a change in the input parameters using the at least one stored function and said determined error; and

updating the current set of input parameters with the determined change in the input parameters.

An exemplary embodiment of the present invention will now be described with reference to the accompanying drawings in which:

Figure 1 is a schematic block diagram illustrating a general arrangement of a computer system which can be programmed to implement the present invention;

Figure 2 is a block diagram of an appearance model generation unit which receives some of the image frames of a source video sequence together with a target image frame and generates therefrom an appearance model;

Figure 3 is a block diagram of a target video sequence generation unit which generates a target video sequence from a source video sequence using a set of stored difference parameters;

Figure 4 is a flow chart illustrating the processing steps which the target video sequence generation unit shown in Figure 3 performs to generate the target video sequence;

Figure 5 schematically illustrates the form of a hierarchical appearance model generated in one embodiment

of the invention;

Figure 6 shows a head with a mesh of triangular facets placed over the head and whose positions are defined by the position of landmark points at the corners of the facets;

Figure 7 is a flow chart illustrating the processing steps required to generate a facet appearance model from the training images;

Figure 8 schematically illustrates the way in which a transformation is defined between a facet in a training image and a predefined shape of facet which allows texture information to be extracted from the facet;

Figure 9 is a flow chart illustrating the main processing steps involved in determining an appearance model for the mouth using the appearance models for the facets which appear in the mouth and using the training images;

Figure 10 schematically illustrates the way in which training images are used to determine some of the appearance models which form the hierarchical appearance model illustrated in Figure 5;

Figure 11a is a flow chart illustrating the processing steps performed during a training routine to identify an

Active matrix associated with a current facet;

Figure 11b is a flow chart illustrating the processing steps performed during a training routine to identify an Active matrix associated with the mouth;

Figure 12 is a flow chart illustrating the processing steps involved in determining a set of parameters which define the appearance of a face within a input image;

Figure 13a shows three frames of an example source video sequence which is applied to the target video sequence generation unit shown in Figure 4;

Figure 13b shows an example target image used to generate a set of difference parameters used by the target video sequence generation unit shown in Figure 4;

Figure 13c shows a corresponding three frames from a target video sequence generated by the target video sequence generation unit shown in Figure 4 from the three frames of the source video sequence shown in Figure 13a using the difference parameters generated using the target image shown in Figure 13b;

Figure 13d shows a second example of a target image used to generate a set of difference parameters for use by the target video sequence generation unit shown in Figure 4;

and

Figure 13e shows the corresponding three frames from the target video sequence generated by the target video sequence generation unit shown in Figure 4 when the three frames of the source video sequence shown in Figure 13a are input to the target video sequence generation unit together with the difference parameters calculated using the target image shown in Figure 13d.

Figure 1 is an image processing apparatus according to an embodiment of the present invention. The apparatus comprises a computer 1 having a central processing unit (CPU) 3 connected to a memory 5 which is operable to store a program defining the sequence of operations of the CPU 3 and to store object and image data used in calculations by the CPU 3. Coupled to an input port of the CPU 3 there is an input device 7, which in this embodiment comprises a keyboard and a computer mouse. Instead of, or in addition to the computer mouse, another position sensitive input device (pointing device) such as a digitiser with associated stylus may be used.

A frame buffer 9 is also provided and is coupled to the CPU 3 and comprises a memory unit (not shown) arranged to store image data relating to at least one image, for example by providing one (or several) memory location(s) per pixel of the image. The value stored in the frame

buffer for each pixel defines the colour or intensity of that pixel in the image. In this embodiment, the images are represented by 2-D arrays of pixels, and are conveniently described in terms of Cartesian coordinates, so that the position of a given pixel can be described by a pair of x-y coordinates. This representation is convenient since the image is displayed on a raster scan display 11. Therefore, the x-coordinate maps to the distance along the line of the display and the y-coordinate maps to the number of the line. The frame buffer 9 has sufficient memory capacity to store at least one image. For example, for an image having a resolution of 1000 x 1000 pixels, the frame buffer 9 includes 10^6 pixel locations, each addressable directly or indirectly in terms of a pixel coordinate x,y.

In this embodiment, a video tape recorder (VTR) 13 is also coupled to the frame buffer 9, for recording the image or sequence of images displayed on the display 11. A mass storage device 15, such as a hard disc drive, having a high data storage capacity is also provided and coupled to the memory 5. Also coupled to the memory 5 is a floppy disc drive 17 which is operable to accept removable data storage media, such as a floppy disc 19 and to transfer data stored thereon to the memory 5. The memory 5 is also coupled to a printer 21 so that generated images can be output in paper form, an image input device 23 such as a scanner or video camera and a

modem 25 so that input images and output images can be received from and transmitted to remote computer terminals via a data network, such as the Internet. The CPU 3, memory 5, frame buffer 9, display unit 11 and mass storage device 13 may be commercially available as a complete system, for example as an IBM compatible personal computer (PC) or a workstation such as the Sparc station available from Sun Microsystems.

A number of embodiments of the invention can be supplied commercially in the form of programs stored on a floppy disc 19 or on other mediums, or as signals transmitted over a data link, such as the Internet, so that the receiving hardware becomes reconfigured into an apparatus embodying the present invention.

In this embodiment, the computer 1 is programmed to receive a source video sequence input by the image input device 23 and to generate a target video sequence from the source video sequence using a target image. In this embodiment, the source video sequence is a video clip of an actor acting out a scene, the target image is an image of a second actor and the resulting target video sequence is a video sequence showing the second actor acting out the scene. The way in which this is achieved will now be briefly described with reference to Figures 2 to 4.

In this embodiment, in order to generate the target video

sequence from the source video sequence, a hierarchical parametric appearance model which models the variability of shape and texture of the head images is used. This appearance model makes use of the fact that some prior knowledge is available about the contents of head images in order to facilitate their modelling. For example, it can be assumed that two frontal images of a human face will each include eyes, a nose and a mouth. In this embodiment, as shown in Figure 2, the hierarchical parametric appearance model 35 is generated by an appearance model generation unit 31 from training images which are stored in an image database 32. In this embodiment, all the training images are colour images having 500 x 500 pixels, with each pixel having a red, green and a blue pixel value. The resulting appearance model 35 is a parameterisation of the appearance of the class of head images defined by the heads in the training images, so that a relatively small number of parameters (for example 50) can describe the detailed (pixel level) appearance of a head image from the class. In particular, the hierarchical appearance model 35 defines a function (F) such that:

$$I = F(\underline{p}) \qquad (1)$$

where $\underline{p}$ is the set of appearance parameters (written in vector notation) which generates, through the hierarchical appearance model (F), the face image I. The

structure of the hierarchical appearance model used in this embodiment will be described later.

Once the hierarchical appearance model 35 has been determined, a target video sequence can be generated from a source video sequence. As shown in Figure 3, the source video sequence is input to a target video sequence generation unit 51 which processes the source video sequence using a set of difference parameters 53 to generate and to output the target video sequence. The difference parameters 53 are determined by subtracting the appearance parameters which are generated for the first actor's head in one of the source video frames, from the appearance parameters which are generated for the second actor's head in the target image. The way in which these appearance parameters are determined for these images will be described later. In order that these difference parameters only represent differences in the general shape and colour texture of the two actors' heads, the pose and facial expression of the first actor's head in the source video frame used should match, as closely as possible, the pose and facial expression of the second actor's head in the target image.

The processing steps required to generate the target video sequence from the source video sequence will now be described in more detail with reference to Figure 4. As shown, in step s1, the appearance parameters (p_s^i) for

the first actor's head in the current video frame (I_s^i) are automatically calculated. The way that this is achieved will be described later. Then, in step s3, the difference parameters (p_{dif}) are added to the appearance parameters for the first actor's head in the current video frame to generate:

$$p_{mod}^{i} = p_{s}^{i} + p_{dif} \qquad (2)$$

The resulting appearance parameters (p_{mod}^i) are then used, in step s5, to regenerate the head for the current target video frame. In particular, the modified appearance parameters are inserted into equation (1) above to regenerate a modified head image which is then composited, in step s7, into the source video frame to generate the corresponding target video frame. A check is then made, in step s9, to determine whether or not there are any more source video frames. If there are, then the processing returns to step s1 where the procedure described above is repeated for the next source video frame. If there are no more source video frames, then the processing ends.

Figure 13 illustrates the results of this animation technique (although showing black and white images and not colour). In particular, Figure 13a shows three frames of the source video sequence, Figure 13b shows the target image (which in this embodiment is computer

generated) and Figure 13c shows the corresponding three frames of the target video sequence obtained in the manner described above. As can be seen, an animated sequence of the computer generated character has been generated from a video clip of a real person and a single image of the computer generated character.

HIERARCHICAL APPEARANCE MODEL

In the systems described by Cootes et al and Blanz et al, the parametric model is created by placing a number of landmark points on a training image and then identifying the same landmark points on the other training images in order to identify how the location of and the pixel values around the landmark points vary within the training images. A principal component analysis is then performed on the matrix which consists of vectors of the landmark points. This PCA yields a set of Eigenvectors which describe the directions of greatest variation along which the landmark points change. Their appearance model includes the linear combination of the Eigenvectors plus parameters for translation, rotation and scaling. This single appearance model relates a compact set of appearance parameters to pixel values.

In this embodiment, rather than having a single appearance model for the object, a hierarchical appearance model comprising several appearance models which model variations in components of the object is

used. For example, in the case of human faces, the hierarchical appearance model may include an appearance model for the mouth, one for the left eye, one for the right eye and one for the nose. Since it may be possible to model various components of the object, the particular hierarchical structure which will be used for a particular object and application must first of all be defined by the system designer.

Figure 5 schematically illustrates the structure of the hierarchical appearance model used in this embodiment. As shown, at the top of the hierarchy there is a general face appearance model 61. Beneath the face appearance model there is a mouth appearance model 63, a left eye appearance model 65, a right eye appearance model 67, a left eyebrow appearance model 69, a rest of left eye appearance model 71, a right eyebrow appearance model 73, a rest of right eye appearance model 75 and, in this embodiment, a facet appearance model for each facet defined in the training images. Figure 6 shows the head of a training image in which the set of landmark points has been placed at the appropriate points on the head. As shown, in this embodiment, there are one hundred and forty-eight triangular areas or facets defined by the positions of the landmark points. Therefore, in this embodiment, there are one hundred and forty-eight facet appearance models 77.

The face appearance model 61 operates to relate a small number of "global" appearance parameters to a further set of appearance parameters, some of which are input to facet appearance models 77, some of which are input to the mouth appearance model 63, some of which are input to the left eye appearance model 65 and the rest of which are input to the right eye appearance model 67. The facet appearance models 77 operate to relate the input parameters received from the appearance model which is above it in the hierarchy into corresponding pixel values for that facet. The mouth appearance model 63 is operable to relate the parameters it receives from the face appearance model 61 into a further set of appearance parameters, respective ones of which are output to the respective facet appearance models 77 for the facets which are associated with the mouth. Similarly, the left and right eye appearance models 65 and 67 operate to relate the parameters it receives from the face appearance model 61 into a further set of appearance parameters, some of which are input to the appropriate eyebrow appearance model and the rest of which are input to the appropriate rest of eye appearance model. These appearance models in turn convert these parameters into parameters for input to the facet appearance models associated with the facets which appear in the left and right eyes respectively. In this way, a small compact set of "global" appearance parameters input to the face appearance model 61 can filter through the hierarchical

structure illustrated in Figure 5 to generate a set of pixel values for all the facets in a head which can then be used to regenerate the image of the head.

The way in which the individual appearance models of this hierarchical appearance model are generated in this embodiment will now be described with reference to Figures 6 to 10.

In this embodiment, each of the training images stored in the image database 32 is labelled with eighty six landmark points. In this embodiment, this is performed manually by the user via the user interface 33. In particular, each training image is displayed on the display 11 and the user places the landmark points over the head in the training image. These points delineate the main features in the head, such as the position of the hairline, neck, eyes, nose, ears and mouth. In order to compare training faces, each landmark point is associated with the same point on each face. In this embodiment, the following landmark points are used:

Landmark Point	Associated Position	Landmark Point	Associated Position
LP_1	Left corner of left eye	LP_{44}	Eye, bottom
LP_2	Right corner of right eye	LP_{45}	Eye, top
LP_3	Chin, bottom	LP_{46}	Eye, bottom
LP_4	Right corner of left eye	LP_{47}	Eyebrow, lower

Landmark Point	Associated Position	Landmark Point	Associated Position
LP_5	Left corner of right eye	LP_{48}	Eyebrow, upper
LP_6	Mouth, left	LP_{49}	Cheek, left
LP_7	Mouth, right	LP_{50}	Cheek, right
LP_8	Nose, bottom	LP_{51}	Eyebrow, lower
LP_9	Nose, between eyes	LP_{52}	Eyebrow, upper
LP_{10}	Upper lip, top	LP_{53}	Eyebrow, lower
LP_{11}	Lower lip, bottom	LP_{54}	Eyebrow, upper
LP_{12}	Neck, left, top	LP_{55}	Eyebrow, lower
LP_{13}	Neck, right, top	LP_{56}	Eyebrow, upper
LP_{14}	Face edge left, level with nose	LP_{57}	Eyebrow, lower
LP_{15}	Face edge	LP_{58}	Eyebrow, upper
LP_{16}	Face edge right, level with nose	LP_{59}	Eyebrow, lower
LP_{17}	Face edge	LP_{60}	Eyebrow, upper
LP_{18}	Top of head	LP_{61}	Eyebrow, lower
LP_{19}	Hair edge	LP_{62}	Lower lip, top
LP_{20}	Hair edge	LP_{63}	Centre forehead
LP_{21}	Hair edge	LP_{64}	Upper lip, top left
LP_{22}	Hair edge	LP_{65}	Upper lip, top right
LP_{23}	Hair edge	LP_{66}	Lower lip, bottom right
LP_{24}	Hair edge	LP_{67}	Lower lip, bottom left
LP_{25}	Hair edge	LP_{68}	Eye, top left
LP_{26}	Hair edge	LP_{69}	Eye, top right
LP_{27}	Hair edge	LP_{70}	Eye, bottom right
LP_{28}	Hair edge	LP_{71}	Eye, bottom left
LP_{29}	Bottom, far left	LP_{72}	Eye, top left
LP_{30}	Bottom, far right	LP_{73}	Eye, top right
LP_{31}	Shoulder	LP_{74}	Eye, bottom right

Landmark Point	Associated Position	Landmark Point	Associated Position
LP_{32}	Shoulder	LP_{75}	Eye, bottom left
LP_{33}	Bottom, left	LP_{76}	Lower lip, top left
LP_{34}	Bottom, middle	LP_{77}	Lower lip, top right
LP_{35}	Bottom, right	LP_{78}	Chin, left
LP_{36}	Left forehead	LP_{79}	Chin, right
LP_{37}	Right forehead	LP_{80}	Neck, left
LP_{38}	Centre, between eyebrows	LP_{81}	Neckline, left
LP_{39}	Nose, left	LP_{82}	Neckline
LP_{40}	Nose, right	LP_{83}	Neckline, right
LP_{41}	Nose edge, left	LP_{84}	Neck, right
LP_{42}	Nose edge, right	LP_{85}	Hair edge
LP_{43}	Eye, top	LP_{86}	Hair edge

The result of the manual placement of the landmark points is a table of landmark points for each training image, which identifies the (x, y) coordinate of each landmark point within the image. As shown in Figure 6, these landmark points are also used to define the location of predetermined triangular facets or areas within the training image.

FACET APPEARANCE MODEL

Figure 7 shows a flow chart illustrating the main processing steps involved in this embodiment in determining a facet appearance model for facet (i). As shown, in step s61, the system determines, for each training image, the apex coordinates of facet (i) and

texture values from within facet (i). In order to sample texture from within the facet at corresponding points within each training facet, a transformation which transforms the facet onto a reference facet is determined. Figure 8 illustrates this transformation. In particular, Figure 8 shows facet f_i^v taken from the V-th training image, which is defined by the landmark points (x_1^v, y_1^v), (x_2^v, y_2^v) and (x_3^v, y_3^v). The transformation (T_i^v) which transforms those coordinates onto coordinates (0,0), (1,0) and (0,1) is determined. In this embodiment, the texture information extracted from each training facet is defined by the regular array of pixels shown in the reference facet. In order to determine the corresponding red, green and blue pixel values in the training image, the inverse transformation $([T_i^v]^{-1})$ is used to transform the pixel locations in the reference facet, into corresponding locations in the training facet, from which the RGB pixel values are determined. In this embodiment, this transformation may not result in an exact correspondence with a single image pixel location since the pixel resolution in the actual facet may be different to the resolution in the reference facet. In this embodiment, the texture information (RGB pixel values) which is determined is obtained by interpolating between the surrounding image RGB pixel values. In this embodiment, there are fifty pixels in the regular array of pixels in the reference facet. Therefore, fifty RGB pixel values are extracted for each

training facet. The texture information for facet (i) from the V-th training image can then be represented by a vector (t^{iv}) of the form:

$$\underline{t}^{iv} = [t_1^{iv}, t_2^{iv}, t_3^{iv} \ldots t_{50}^{iv}]^T$$

where t_1^{iv} is the RGB texture information for the first reference pixel extracted from facet (i) in the V-th training image etc.

In this embodiment, the facet appearance models 77 treat shape and texture separately. Therefore, in step s63, the system performs a principal component analysis (PCA) on the set of texture training vectors generated in step s61. For a more detailed discussion of principal component analysis, the reader is referred to the book by W. J. Krzanowski entitled "Principles of Multivariate Analysis - A User's Perspective" 1998, Oxford Statistical Science Series. As those skilled in the art will appreciate, this principal component analysis determines all possible modes of variation within the training texture vectors. However, since each of the facets is associated with a similar point on the face, most of the variation within the data can be explained by a few modes of variation. The result of the principal component analysis is a facet texture appearance model (defined by matrix F_i) which relates a vector of facet texture parameters to a vector of texture pixel values, by:

$$\underline{p}_v^{fit} = F_i(\underline{t}^{iv} - \overline{\underline{t}}^i) \quad (3)$$

where $\underline{t}^{iv}$ is the RGB texture vector defined above, $\overline{\underline{t}}^i$ is the mean RGB texture vector for facet (i), F_i is a matrix which defines the facet texture appearance model for facet (i) and $\underline{p}^{fit}{}_v$ is a vector of the facet texture parameters which describes the RGB texture vector $\underline{t}^{iv}$. The matrix F_i describes the main modes of variation of the texture within the training facets; and the vector of facet texture parameters ($\underline{p}^{fit}{}_v$) for a given input facet has a parameter associated with each mode of variation whose value relates the texture of the input facet to the corresponding mode of variation.

As those skilled in the art will appreciate, for facets which describe fairly constant parts of the face, such as the chin or cheeks, very few parameters will be needed to model the variability within the training images. However, facets which are associated with areas of the face where there is a large amount of variability (such as facets which form part of the eye), will require a larger number of facet texture parameters to describe the variability within the training images. Therefore, in step s65, the system determines how many texture parameters are needed for the current facet and stores the appropriate facet appearance model matrix.

In addition to being able to determine a set of texture

parameters $\underline{p}^{Fit}_{v}$ for a given texture vector $\underline{t}^{iv}$, equation (3) can be solved with respect to the texture vector $\underline{t}^{iv}$ to give:

$$\underline{t}^{iv} = \overline{\underline{t}}^{i} - F_i^T \underline{p}_v^{Fit} \qquad (4)$$

since $F_i F_i^T$ equals the identity matrix. Therefore, by modifying the set of texture parameters ($\underline{p}^{Fit}$) within suitable limits, new textures for facet (i) can be generated which are similar to those in the training set.

Once the above procedure has been performed for each of the one hundred and forty-eight facets in the training images, a facet texture appearance model will have been generated for each of those facets. In this embodiment, the facet appearance model does not compress the parameters defining the shape of the facets, since only six parameters are needed to define the shape of each facet - two parameters for each (x,y) coordinate of the facet's apexes.

MOUTH APPEARANCE MODEL

Figure 9 shows a flow chart illustrating the main processing steps required in order to generate the mouth appearance model 63. As shown, in step s67, the system uses the facet appearance models for the facets which form part of the mouth to generate shape and texture parameters from those facets for each training image.

Therefore, referring to Figure 10, the mouth appearance model 63 will receive texture and shape parameters from the facet appearance model for facet (i), facet (j) and facet (n) for the corresponding facets in each of the training images 79. As illustrated in Figure 10, the appearance model for facet (i) is operable to generate, for each training image, six shape parameters (corresponding to the three (x,y) coordinates of the apexes of facet (i)) and six texture parameters. Similarly, the appearance model for facet (j) is operable to generate, for each training image, six shape parameters and four texture parameters and the appearance model for facet (n) is operable to generate, for each training image, six shape parameters and three texture parameters.

The processing then proceeds to step s69 where the system performs a principal component analysis on the shape and texture parameters generated for the training images by the facet appearance models associated with the mouth. In this embodiment, the mouth appearance model 63 treats the shape and texture separately. In particular, for each training image, the system concatenates the six shape parameters for the facets associated with the mouth to form the following shape vector:

$$\underline{p}^{mouth} = [x_1^{fi}, y_1^{fi}, x_2^{fi}, y_2^{fi}, x_3^{fi}, y_3^{fi} \; : \; x_1^{fj}, y_1^{fj}, x_2^{fj}, y_2^{fj} \; . \; . \; .]^T$$

and concatenates the facet texture parameters output by the facet appearance models associated with the mouth to form the following texture vector:

$$\underline{p}^{Fmt} = [p_1^{Fit}, p_2^{Fit} \ldots p_6^{Fit} : p_1^{Fjt}, p_2^{Fjt} \ldots p_4^{Fjt} : p_1^{Fnt}, \ldots]^T$$

The system then performs a principal component analysis on the shape vectors generated by all the training images to generate a shape appearance model for the mouth (defined by matrix M_s) which relates each mouth shape vector to a corresponding vector of shape mouth parameters by:

$$\underline{p}_v^{Ms} = M_s(\underline{p}_v^{FMs} - \overline{\underline{p}}^{FMs}) \qquad (5)$$

where $\underline{p}^{FMs}{}_v$ is the mouth shape vector for the mouth in the V-th training image, $\overline{\underline{p}}^{FMs}$ is the mean mouth shape vector from the training vectors and $\underline{p}^{Ms}{}_v$ is a vector of mouth shape parameters for the mouth shape vector $\underline{p}^{FMs}{}_v$. The mouth shape model, defined by matrix M_s, describes the main modes of variation of the shape of the mouths within the training images; and the vector of mouth shape parameters ($\underline{p}^{Ms}{}_v$) for the mouth in the V-th training image has a parameter associated with each mode of variation whose value relates the shape of the input mouth to the corresponding mode of variation.

As with the facet appearance models, equation (5) above

can be rewritten with respect to the mouth shape vector p^{FMs}_{v} to give:

$$p_v^{FMs} = \bar{p}^{FMs} - M_s^T p_v^{Ms} \quad (6)$$

since M_s M_s^T equals the identity matrix. Therefore, by modifying the mouth shape parameters, new mouth shapes can be generated which will be similar to those in the training set.

The system then performs a principal component analysis on the mouth texture parameter vectors (p^{FMt}) which are generated for the training images. This principal component analysis generates a mouth texture model (defined by matrix M_t) which relates each of the facet texture parameter vectors for the facets associated with the mouth, to a corresponding vector of mouth texture parameters, by:

$$p_v^{Mt} = M_t(p_v^{FMt} - \bar{p}^{FMt}) \quad (7)$$

where p^{FMt}_{v} is a vector of mouth facet texture parameters generated by the facet appearance models associated with the mouth for the mouth in the V-th training image; $\bar{p}^{FMt}$ is the mean vector of mouth facet texture parameters from the training vectors and p^{Mt}_{v} is a vector of mouth texture parameters for the facet texture parameters p^{FMt}_{v}. The matrix M_t describes the main modes of variation within

the training images of the facet texture parameters generated by the facet appearance models which are associated with the mouth; and the vector of mouth texture parameters ($\underline{p}^{MT}_{v}$) has a parameter associated with each of those modes of variation whose value relates the texture of the input mouth to the corresponding mode of variation.

The processing then proceeds to step s71 shown in Figure 9 where the system determines the number of shape parameters and texture parameters needed to describe the training data received from the facet appearance models which are associated with the mouth. As shown in Figure 10, in this embodiment, the mouth appearance model 63 requires five shape parameters and four texture parameters to be able to model most of this variation. The system therefore stores the appropriate mouth shape and texture appearance model matrices for subsequent use.

As those skilled in the art will appreciate, a similar procedure is performed to determine each of the appearance models shown in Figure 5, starting from the facet appearance models at the base of the hierarchy. A further description of how these remaining appearance models are determined will, therefore, not be given here. The resulting hierarchical appearance model allows a small number of global face appearance parameters to be input to the face appearance model 61, which generates

further parameters which propagate down through the hierarchical model structure until facet pixel values are generated, from which an image which corresponds to the global appearance parameters can be generated.

AUTOMATIC GENERATION OF APPEARANCE PARAMETERS

In the description given above of the way in which the appearance models are generated, appearance parameters for an image were generated from a manual placement of a number of landmark points over the image. However, during use of the appearance model to track the first actor's head in the source video sequence and during the calculation of the difference parameters (p_{dif}), the appearance parameters for the heads in the input images were automatically calculated. This task involves finding the set of global appearance parameters p which best describe the pixels in view. This problem is complicated because the inverse of each of the appearance models in the hierarchical appearance model is not necessarily one-to-one. In this embodiment, the appearance parameters for the head in an input image are calculated in a two-step process. In the first step, an initial set of global appearance parameters for the head in the current frame (I_s^i) is found using a simple and rapid technique. For all but the first frame of the source video sequence, this is achieved by simply using the appearance parameters from the preceding video frame (I_s^{i-1}) before modification in step s3 (i.e. parameters

$\underline{p}_s^{i-1}$). In this embodiment, the global appearance parameters ($\underline{p}$) effectively define the shape and colour texture of the head. For the first frame and for the target image the initial estimate of the appearance parameters is set to the mean set of appearance parameters and the scale, position and orientation is initially estimated by the user manually placing the mean head over the head in the image.

In the second step, an iterative technique is used in order to make fine adjustments to the initial estimate of the appearance parameters. The adjustments are made in an attempt to minimise the difference between the head described by the global appearance parameters (the model head) and the head in the current video frame (the image head). With 50 appearance parameters, this represents a difficult optimisation problem. This can be performed by using a standard steepest descent optimisation technique to iteratively reduce the mean squared error between the given image pixels and those predicted by a particular set of appearance parameter values. In particular, minimising the following error function $E(\underline{p})$:

$$E(\underline{p}) = [I^a - F(\underline{p})]^T [I^a - F(\underline{p})] \quad (8)$$

where I^a is a vector of actual image RGB pixel values at the locations where the appearance model predicts values

(the appearance model does not predict all pixel values since it ignores background pixels and only predicts a subsample of pixel values within the object being modelled) and $F(\underline{p})$ is the vector of image RGB pixel values predicted by the hierarchical appearance model. As those skilled in the art will appreciate, $E(\underline{p})$ will only be zero when the model head (i.e. $F(\underline{p})$) predicts the actual image head (I^a) exactly. Standard steepest descent optimisation techniques stipulate that a step in the direction $-\nabla E(\underline{p})$ should result in a reduction in the error function $E(\underline{p})$, provided the error function is well behaved. Therefore, the change ($\Delta\underline{p}$) in the set of parameter values should be:

$$\Delta\underline{p} = 2[\nabla F(\underline{p})]^T[I^a - F(\underline{p})] \quad (9)$$

which requires the calculation of the differential of the appearance model, i.e. $\nabla F(\underline{p})$.

The technique described by Edwards et al assumes that, on average over the whole parameter space, $\nabla F(\underline{p})$ is constant. The update equation then becomes:

$$\Delta\underline{p} = A[I^a - F(\underline{p})] \quad (10)$$

for some constant matrix A (referred to as the "Active matrix") which is determined beforehand during a training routine. In this embodiment, rather than using a single

constant matrix associated with the entire hierarchical appearance model, an Active matrix is determined and used for each of the individual appearance models which form part of the hierarchical appearance model. The way in which these Active matrices are determined in this embodiment will now be described with reference to Figures 11a and 11b, which illustrate the processing steps performed to generate the Active matrix for each facet appearance model and the Active matrix for the mouth appearance model.

As shown in Figure 11a, in step s73, the system chooses a random facet parameter vector ($\underline{p}^{Fi}$) for the current facet (i) and then, in s75, perturbs this facet parameter vector by a small random amount to create $\underline{p}^{Fi} + \Delta\underline{p}^{Fi}$. In this embodiment, the facet parameter vectors include not only the texture parameters, but also the six shape parameters which define the (x,y) coordinates of the facet's location within the image. The processing then proceeds to step s77 where the system uses the parameter vector $\underline{p}^{Fi}$ and the perturbed parameter vector $\underline{p}^{Fi} + \Delta\underline{p}^{Fi}$ to create model images I_0^{Fi} and I_1^{Fi} respectively. The processing then proceeds to step s79 where the system records the parameter change $\Delta\underline{p}^{Fi}$ and image difference $I_1^{Fi} - I_0^{Fi}$. Then in step s81, the system determines whether or not there is sufficient training data for the current facet. If there is not then the processing returns to step s21. Once sufficient training data has been

generated, the processing proceeds to step s83 where the system performs multiple multivariate linear regressions on the data for the current facet to identify an Active matrix (A_{Fi}) for the current facet.

Figure 11b shows the processing steps required to calculate the Active matrix for the mouth appearance model. As shown, in step s85, the system chooses a random mouth parameter vector $\underline{p}^M$. In this embodiment, this vector includes both the mouth shape parameters and the mouth texture parameters. Then, in step s87, the system perturbs this mouth parameter vector by a small random amount to create $\underline{p}^M + \Delta\underline{p}^M$. The processing then proceeds to step s89 where the system uses the mouth parameter vectors $\underline{p}^M$ and the perturbed mouth parameter vector $\underline{p}^M + \Delta\underline{p}^M$ to create model images I_0^m and I_1^m respectively, using the mouth appearance model and the facet appearance models associated with the mouth. The processing then proceeds to step s91 where the facet appearance models associated with the mouth are used again to transform the mouth model images I_0^m and I_1^m into corresponding facet appearance parameters $\underline{p}_0^{FM}$ and $\underline{p}_1^{FM}$, which are then subtracted to determine the corresponding change $\Delta\underline{p}^{FM}$ in the mouth facet parameters. The processing then proceeds to step s93 where the system records the mouth parameter change $\Delta\underline{p}^M$ and the mouth facet parameter change $\Delta\underline{p}^{FM}$. The processing then proceeds to step s95 where the system determines whether

or not there is sufficient training data. If there is not, then the processing returns to step s85. Once sufficient training data has been generated, the processing proceeds to step s97, where the system performs multiple multivariate linear regressions on the training data for the mouth to identify the Active matrix (A_M) for the mouth which relates changes in mouth parameters Δp^M to changes in facet parameters Δp^{FM} for the facets associated with the mouth.

As those skilled in the art will appreciate, a similar processing technique is used in order to identify the Active matrix for each of the appearance models shown in Figure 5.

Once the Active matrices have been determined for the hierarchical appearance model, they can then be used to iteratively update a current estimate of a set of appearance parameters for an input image. Figure 12 illustrates the processing steps performed in this iterative routine for the current source video frame. As shown, in step s101, the system initially estimates a set of global parameters for the head in the current source video frame. The processing then proceeds to step s103 where the system generates a model image from the estimated global parameters and the hierarchical appearance model. The system then proceeds to step s105 where it determines the image error between the model

image and the current source video frame. Then, in step s107, the system uses this image error to propagate parameter changes up the hierarchy of the hierarchical appearance model using the stored Active matrices to determine a change in the global parameters. This change in global parameters is then used, in step s109, to update the current global parameters for the current source video frame. The system then determines, in step s111, whether or not convergence has been reached by comparing the error obtained from equation (8) using the updated global parameters with a predetermined threshold (Th). If convergence has not been reached, then the processing returns to step s103. Once convergence is reached, the processing proceeds to step s113, where the current global appearance parameters are output as the global appearance parameters for the current source video frame and then the processing ends.

ALTERNATIVE EMBODIMENTS

In the above embodiment, the same hierarchical model structure was used to model the variation in the shape and texture within the training images. As those skilled in the art will appreciate, one model hierarchy can be used to model the shape variation and a different model hierarchy can be used to model the texture variation. Alternatively still, rather than separating the shape and texture parameters, each of the appearance models within the hierarchical model may model the combined variation

of the shape and texture within the training images.

In the above embodiments, a facet appearance model was generated for each facet defined within the training images. As those skilled in the art will appreciate, many of the facets may be grouped together such that a single facet appearance model is generated for those facets. In one form of such an embodiment, a single facet appearance model may be determined which models the variability of texture within each facet of the training images.

In the above embodiments, the same amount of texture information was extracted from each facet within the training images. In particular, fifty RGB texture values were extracted from each training facet. In an alternative embodiment, the amount of texture information extracted from each facet may vary in dependence upon the size of the facet. For example, more texture information may be extracted from larger facets or more texture information may be extracted from facets associated with important features of the face, such as the mouth, eyes or nose.

In the above embodiments, each appearance model was determined from a principal component analysis of a set of training data. This principal component analysis determines a linear relationship between the training

data and a set of model parameters. As those skilled in the art will appreciate, techniques other than principal component analysis can be used to determine a parametric model which relates a set of parameters to the training data. This model may define a non-linear relationship between the training data and the model parameters. For example, one or more of the models within the hierarchy may comprise a neural network which relates the set of input parameters to the training data.

In the above embodiments, a principal component analysis was performed on a set of training data in order to identify a relatively small number of parameters which describe the main modes of variation within the training data. This allows a relatively small number of input parameters to be able to generate a larger set of output parameters from the model. However, as those skilled in the art will appreciate, this is not essential. One or more of the appearance models may act as transformation models in which the number input parameters is the same as or greater than the number of output parameters. This can be used to generate a set of input parameters which can be changed by the user in some intuitive way. For example, in order to identify parameters which have a linear relationship with features in the object, such as a parameter that linearly changes the amount of smile within a face image.

In the above embodiments, a set of Active matrices were used in order to identify automatically a set of appearance parameters for an input image. As those skilled in the art will appreciate, rather than having separate Active matrices for each of the components in the hierarchical appearance model, a global Active matrix may be used instead. Further, although both the shape and grey level parameters were used in order to derive the Active matrices, suitable Active matrices can be determined using just the shape information.

In the above embodiments, the variation in both the shape and texture within the training images were modelled. As those skilled in the art will appreciate, this hierarchical modelling technique can be used to model only the shape of the objects within the training images. Such a shape model could be then used to track objects within a video sequence.

In the first embodiment, the target image illustrated a computer generated head. This is not essential. For example, the target image might be a hand-drawn head or an image of a real person. Figures 13d and 13e illustrate how an embodiment with a hand-drawn character might be used in character animation. In particular, Figure 13d shows a hand-drawn sketch of a character which, when combined with the images from the source video sequence (some of which are shown in Figure 13a)

generate a target video sequence, some frames of which are shown in Figure 13e. As can be seen from a comparison of the corresponding frames in the source and target video frames, the hand-drawn sketch has been animated automatically using this technique. As those skilled in the art will appreciate, this is a much quicker and simpler technique for achieving computer animation as compared with existing systems which require the animator to manually create each frame of the animation. In particular, in this embodiment, all that is required is a video sequence of a real life actor acting out the scene to be animated, together with a single sketch of the character to be animated.

The above embodiment has described the way in which a target image can be used to modify a source video sequence. In order to do this, a set of appearance parameters has to be automatically calculated for each frame in the video sequence. This involved the use of a number of Active matrices which relate image errors to appearance parameter changes. As those skilled in the art will appreciate, similar processing is required in other applications, such as the tracking of an object within a video sequence, the tracking of a human face within a video sequence or the tracking of a knee joint in an MRI scan.

In the above embodiment, the appearance model was used to

model the variations in facial expressions and 3D pose of human heads. As those skilled in the art will appreciate, the appearance model can be used to model the appearance of any deformable object such as parts of the body and other animals and objects. For example, the above techniques can be used to track the movement of lips in a video sequence. Such an embodiment could be used in film dubbing applications in order to synchronise the lip movements with the dubbed sound. This animation technique might also be used to give animals and other objects human-like characteristics by combining images of them with a video sequence of an actor. This technique can also be used for monitoring the shape and appearance of objects passing along a production line for quality control purposes.

In the above embodiment, the appearance model was generated by using a principal component analysis of shape and texture data which is extracted from the training images. As those skilled in the art will appreciate, by modelling the features of the training heads in this way, it is possible to accurately model each head by just a small number of parameters. However, other modelling techniques, such as vector quantisation and wavelet techniques can be used.

In the above embodiments, the training images used to generate the appearance model were all colour images in

which each pixel had an RGB value. As those skilled in the art will appreciate, the way in which the colour is represented in this embodiment is not important. In particular, rather than each pixel having a red, green and blue value, they might be represented by a chrominance and a luminance component or by hue, saturation and value components. Alternatively still, the training images may be black and white images, in which case only grey level data would be extracted from the facets in the training images. Additionally, the resolution of each training image may be different.

In the above embodiment, during the automatic generation of the appearance parameters, and in particular during the iterative updating of these appearance parameters the error between the input image and the model image was generated using the appearance model. Since this iterative technique still requires a relatively accurate initial estimate for the appearance parameters, it is possible initially to perform the iterations using lower resolution images and once convergence has been reached for the lower resolutions to then increase the resolution of the images and to repeat the iterations for the higher resolutions. In such an embodiment, separate Active matrices would be required for each of the resolutions.

In the above embodiment, the difference parameters were determined by comparing the image of the first actor from

one of the frames of the source video sequence with the image of the second actor in the target image. In an alternative embodiment, a separate image of the first actor may be provided which does not form part of the source video sequence.

In the above embodiments, each of the appearance models modelled variations in two-dimensional images. The above modelling technique could be adapted to work with 3D images and animations. In such an embodiment, the training images used to generate the appearance model would normally include 3D images instead of 2D images. The three-dimensional models may be obtained using a three dimensional scanner which typically work either by using laser range-finding over the object or by using one or more stereo pairs of cameras. Once a 3D hierarchical appearance model has been created from the training models, new 3D models can be generated by adjusting the appearance parameters and existing 3D models can be animated using the same differencing technique that was used in the two-dimensional embodiment described above. This 3D model can then be used to track 3D objects directly within a 3D animation. Alternatively, a 2D model may be used to track the 3D object within a video sequence and then use the result to generate 3D data for the tracked object.

In the above embodiment, a set of difference parameters

were identified which describe the main differences between the head in the video sequence and the head in the target image, which difference parameters were used to modify the video sequence so as to generate a target video sequence showing the second head. In the embodiment, the set of difference parameters were added to a set of appearance parameters for the current frame being processed. In an alternative embodiment, the difference parameters may be weighted so that, for example, the target video sequence shows a head having characteristics from both the first and second actors.

In the above embodiment, a hierarchical appearance model is used to model the appearance of human faces. The model is then used to modify a source video sequence showing a first actor performing a scene to generate a target video sequence showing a second actor performing the same scene. As those skilled in the art will appreciate, the hierarchical model presented above can be used in various other applications. For example, the hierarchical appearance model can be used for synthetic two-dimensional or three-dimensional character generation; video compression when the video is substantially that of an object which is modelled by the appearance model; object recognition for security purposes; face tracking for human performance analysis or human computer interaction and the like; 3D model generation from two-dimensional images; and image editing

(for example making people look older or younger, fatter or thinner etc).

In the above embodiment, an iterative process was used to update an estimated set of appearance parameters for an input image. This iterative process continued until an error between the actual image and the image predicted by the model was below a predetermined threshold. In an alternative embodiment, where there is only a predetermined amount of time available for determining a set of appearance parameters for an input image, this iterative routine may be performed for a predetermined period of time or for a predetermined number of iterations.

CLAIMS:

1. A parametric model for modelling the shape of an object, the model comprising:

data defining a function which relates a set of input parameters to a set of locations which identify the relative positions of a plurality of predetermined points on the object;

characterised in that said data defines a hierarchical set of functions in which a function in a top layer of the hierarchy is operable to generate a set of output parameters from a set of input parameters and in which one or more functions in a bottom layer of the hierarchy are operable to receive parameters output from one or more functions from a higher layer of the hierarchy and to generate therefrom at least some of said locations which identify the relative positions of said predetermined points.

2. A model according to claim 1, wherein said hierarchy comprises one or more intermediate layers of functions which are operable to receive parameters output from one or more functions from a higher layer of the hierarchy and to generate therefrom a set of output parameters for input to functions in a lower layer of the hierarchy.

3. A model according to claim 1 or 2, for modelling the two-dimensional shape of the object by identifying the

relative positions of said predetermined points in a predetermined plane.

4. A model according to claim 1 or 2, for modelling the three-dimensional shape of the object by identifying the relative positions of the predetermined points in a three-dimensional space.

5. A model according to any preceding claim, wherein one or more of said functions comprises a linear function which linearly relates the input parameters to the function to the output parameters of the function.

6. A model according to claim 5, wherein said one or more linear functions are identified from a principal component analysis of training data derived from a set of training objects.

7. A model according to any preceding claim, wherein one or more of said functions are non-linear.

8. A model according to claim 7, wherein at least one of said non-linear functions comprises a neural network.

9. A model according to any preceding claim, wherein the number of parameters input to at least one of said functions is smaller than the number of parameters output from the function.

10. A model according to any preceding claim, wherein the number of input parameters to at least one of said functions is greater than or equal to the number of parameters output by the function.

11. A model according to any preceding claim for modelling the shape and texture of the object, the model further comprising data defining a hierarchical set of functions in which a function in a top layer of the hierarchy is operable to generate a set of output parameters from a set of input parameters and in which one or more functions in a bottom layer of the hierarchy are operable to receive parameters output from one or more functions from a higher layer of the hierarchy and to generate therefrom texture information for the object.

12. A model according to claim 11, wherein the texture hierarchy has the same structure as the shape hierarchy.

13. A model according to claim 11 or 12, wherein one or more of said functions are operable to relate an input set of shape and texture parameters to an output set of appearance parameters defining both shape and texture.

14. A model according to any preceding claim, wherein said object is a deformable object.

15. A model according to claim 14, wherein said

deformable object includes a human face.

16. A model according to claim 15, wherein said function in said top layer of the hierarchy models the shape of the entire face and wherein said hierarchy includes a function which models the shape of the mouth.

17. A model according to claim 16, wherein said hierarchy further comprises a function for modelling the shape of the eyes.

18. A model according to any preceding claim, wherein the or each function in the bottom layer of the hierarchy identifies the positions of a plurality of predetermined points according to a predefined function of smaller number of control point positions.

19. A model according to claim 18, wherein the predefined function for each of the plurality of points is a linear mapping of the control point positions and the control points are the three corners of a triangular facet.

20. A model according to claim 18, wherein the predefined function for each of the plurality of points is a predefined non-linear mapping of a fixed number of control point positions.

21. A model according to claim 18, wherein the predefined function for each of the plurality of points is a predefined displacement from a single control point.

22. A method of determining a set of appearance parameters representative of the appearance of an object, the method comprising the steps of:

(i) storing a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object;

(ii) storing at least one function which relates a change in the input parameters to an error between actual appearance data for the object and appearance data determined from the set of input parameters and said parametric model;

(iii) initially estimating a current set of input parameters for the object;

(iv) determining appearance data for the object from the current set of input parameters and the stored parametric model;

(v) determining the error between actual appearance data of the object and the appearance data determined from the current set of input parameters;

(vi) determining a change in the input parameters using said at least one stored function and said determined error; and

(vii) updating the current set of input parameters

with the determined change in the input parameters.

23. A method according to claim 22, further comprising the step of repeating steps (iv) to (vii) until the error determined in step (v) is less than a predetermined threshold.

24. A method according to claim 22, further comprising the step of repeating steps (iv) to (vii) for a predetermined amount of time or for a predetermined number of repetitions.

25. A method according to claim 22, 23 or 24, wherein said second storing step stores a plurality of functions, one associated with each function within the hierarchical model.

26. A method of tracking an object comprising the steps of:

(i) storing a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object;

(ii) storing at least one function which relates a change in the input parameters to an error between the actual appearance data for the object and the appearance data determined from the set of input parameters and said parametric model;

(iii) initially estimating a current set of input parameters for the object;

(iv) determining the appearance data for the object from the current set of input parameters and the stored parametric model;

(v) determining an error between the actual appearance data for the object and the appearance data for the object determined from the current set of input parameters;

(vi) determining a change in the input parameters using the at least one stored function and the determined error;

(vii) updating the current set of input parameters with said change in the input parameters;

(viii) repeating steps (iv) to (vii) in order to reduce the error determined in step (v); and

(ix) repeating steps (iii) to (viii) to track the object.

27. An apparatus for determining a set of appearance parameters representative of the appearance of an object, the apparatus comprising:

means for storing (i) a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object; and (ii) at least one function which relates a change in the input parameters to an error between actual appearance data for the object and

the appearance data for the object determined from the set of input parameters and said parametric model;

means for receiving an initial estimate of a current set of input parameters for the object;

means for updating the current set of input parameters comprising:

(i) means for determining appearance data for the object from the current set of input parameters and the stored parametric model;

(ii) means for determining the error between the actual appearance data for the object and the appearance data for the object determined from the current set of input parameters;

(iii) means for determining a change in the input parameters using said at least one stored function and said determined error; and

(iv) means for updating the current set of input parameters with the determined change in the input parameters.

28. An apparatus according to claim 27, wherein said updating means is operable to update iteratively the current set of input parameters until the error determining means determines an error which is less than a predetermined threshold.

29. An apparatus according to claim 27 or 28, wherein said storing means stores a plurality of functions, one

associated with each function within the hierarchical model.

30. An apparatus for tracking an object comprising:

means for storing (i) a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object; and (ii) at least one function which relates a change in the input parameters to an error between actual appearance data for the object and the appearance data for the object determined from the set of input parameters and said parametric model;

means for receiving an initial estimate of a current set of input parameters for the object;

means for updating the current set of input parameters comprising:

(i) means for determining appearance data for the object from the current set of input parameters and the stored parametric model;

(ii) means for determining an error between actual appearance data for the object and the appearance data for the object determined from the current set of input parameters;

(iii) means for determining a change in the input parameters using the at least one stored function and the determined error; and

(iv) means for updating the current set of input parameters with said change in the input parameters;

wherein said updating means is operable to update iteratively the current set of input parameters in order to reduce the determined error, wherein said receiving means is operable to receive further estimates of the current input parameters and wherein said update means is operable to update the received estimates of the current input parameters in order to track said object.

31. A storage medium storing the parametric model according to any of claims 1 to 21 or storing processor implementable instructions for controlling a processor to implement the method of any one of claims 22 to 26.

32. Processor implementable instructions for controlling a processor to implement the method of any one of claims 22 to 26.

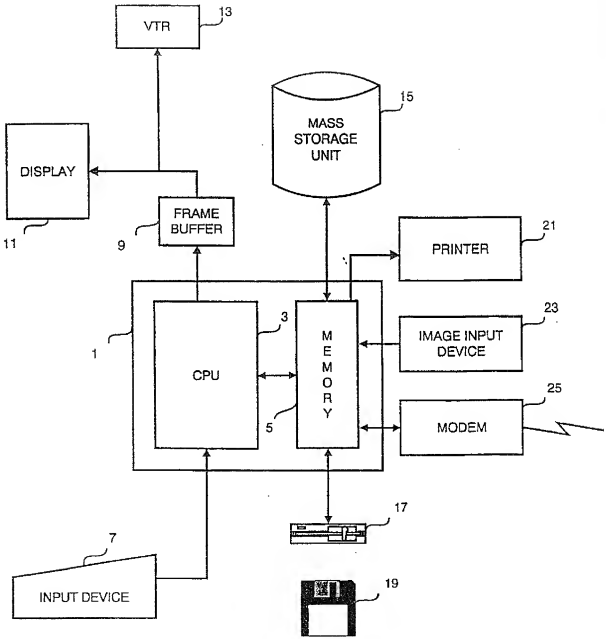

VTR
13
MASS STORAGE UNIT
15
DISPLAY
11
FRAME BUFFER
9
PRINTER
21
1
CPU
3
M E M O R Y
5
IMAGE INPUT DEVICE
23
MODEM
25
17
7
INPUT DEVICE
19

Fig. 1

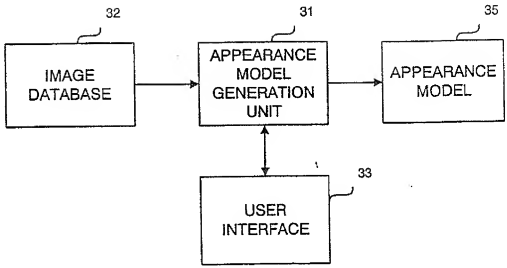

32
31
35
IMAGE DATABASE
APPEARANCE MODEL GENERATION UNIT
APPEARANCE MODEL
33
USER INTERFACE

Fig. 2

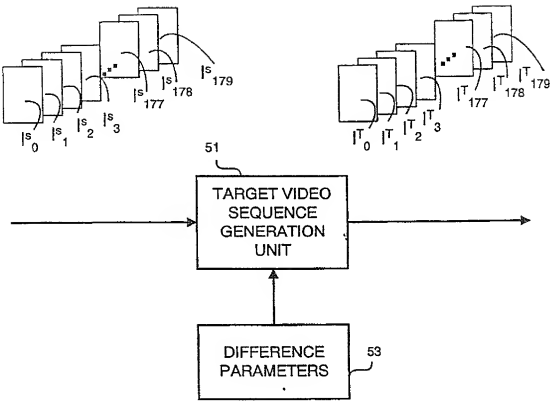

I^S_0 I^S_1 I^S_2 I^S_3 I^S_{177} I^S_{178} I^S_{179}
I^T_0 I^T_1 I^T_2 I^T_3 I^T_{177} I^T_{178} I^T_{179}
51
TARGET VIDEO SEQUENCE GENERATION UNIT
DIFFERENCE PARAMETERS
53

Fig. 3

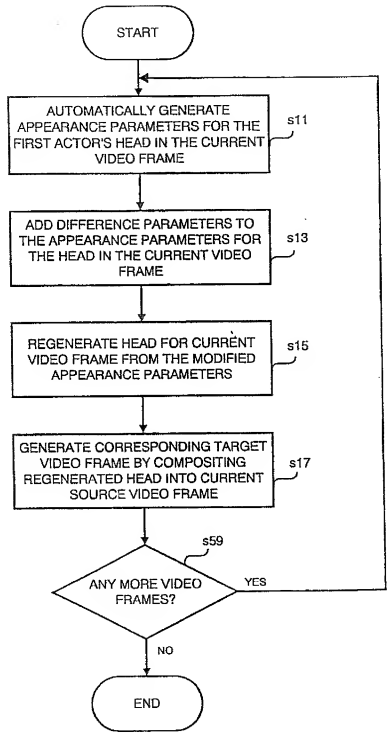

START
AUTOMATICALLY GENERATE APPEARANCE PARAMETERS FOR THE FIRST ACTOR'S HEAD IN THE CURRENT VIDEO FRAME
s11
ADD DIFFERENCE PARAMETERS TO THE APPEARANCE PARAMETERS FOR THE HEAD IN THE CURRENT VIDEO FRAME
s13
REGENERATE HEAD FOR CURRENT VIDEO FRAME FROM THE MODIFIED APPEARANCE PARAMETERS
s15
GENERATE CORRESPONDING TARGET VIDEO FRAME BY COMPOSITING REGENERATED HEAD INTO CURRENT SOURCE VIDEO FRAME
s17
s59
ANY MORE VIDEO FRAMES?
YES
NO
END

Fig. 4

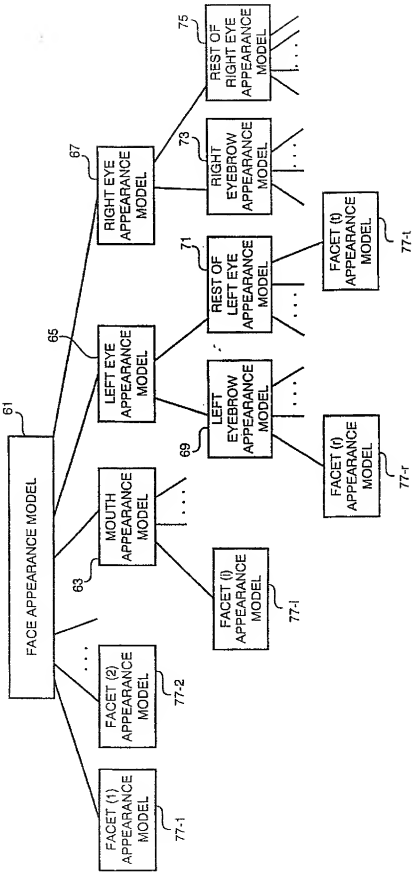

FACE APPEARANCE MODEL
61
FACET (1) APPEARANCE MODEL
77-1
FACET (2) APPEARANCE MODEL
77-2
FACET (i) APPEARANCE MODEL
77-i
FACET (f) APPEARANCE MODEL
77-f
FACET (t) APPEARANCE MODEL
77-t
MOUTH APPEARANCE MODEL
63
LEFT EYEBROW APPEARANCE MODEL
69
LEFT EYE APPEARANCE MODEL
65
REST OF LEFT EYE APPEARANCE MODEL
71
RIGHT EYE APPEARANCE MODEL
67
RIGHT EYEBROW APPEARANCE MODEL
73
REST OF RIGHT EYE APPEARANCE MODEL
75

Fig. 5

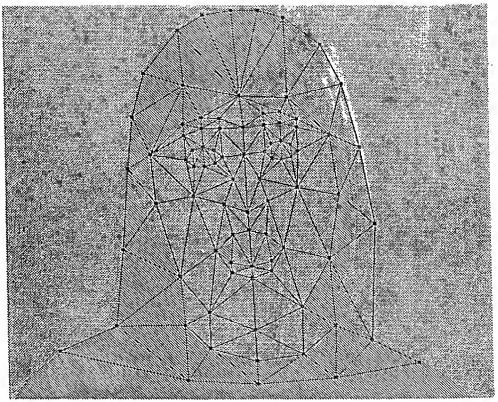


Fig. 6

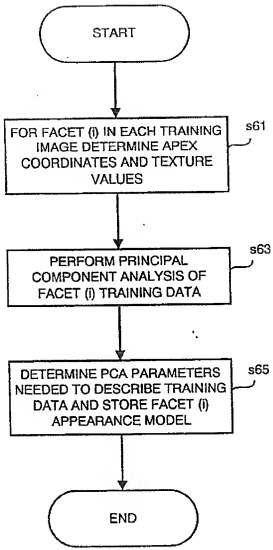

START
FOR FACET (i) IN EACH TRAINING IMAGE DETERMINE APEX COORDINATES AND TEXTURE VALUES
s61
PERFORM PRINCIPAL COMPONENT ANALYSIS OF FACET (i) TRAINING DATA
s63
DETERMINE PCA PARAMETERS NEEDED TO DESCRIBE TRAINING DATA AND STORE FACET (i) APPEARANCE MODEL
s65
END

Fig. 7

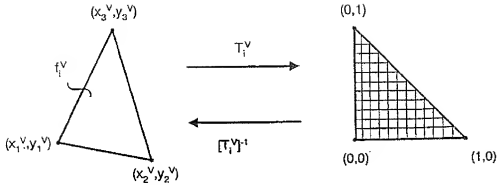


Fig. 8

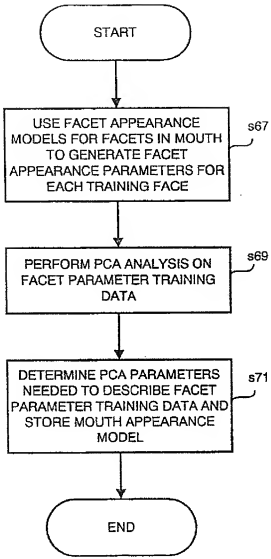
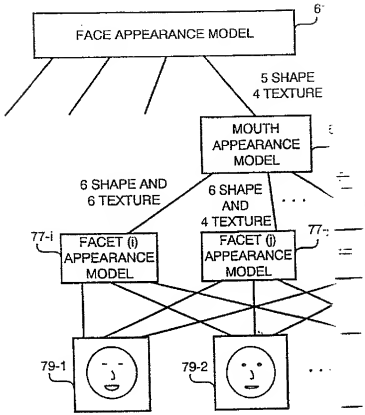

START
USE FACET APPEARANCE MODELS FOR FACETS IN MOUTH TO GENERATE FACET APPEARANCE PARAMETERS FOR EACH TRAINING FACE
s67
PERFORM PCA ANALYSIS ON FACET PARAMETER TRAINING DATA
s69
DETERMINE PCA PARAMETERS NEEDED TO DESCRIBE FACET PARAMETER TRAINING DATA AND STORE MOUTH APPEARANCE MODEL
s71
END

Fig. 9


FACE APPEARANCE MODEL
6
5 SHAPE
4 TEXTURE
MOUTH
APPEARANCE
MODEL
6 SHAPE AND
6 TEXTURE
6 SHAPE
AND
4 TEXTURE
77-i
FACET (i)
APPEARANCE
MODEL
FACET (j)
APPEARANCE
MODEL
79-1
79-2

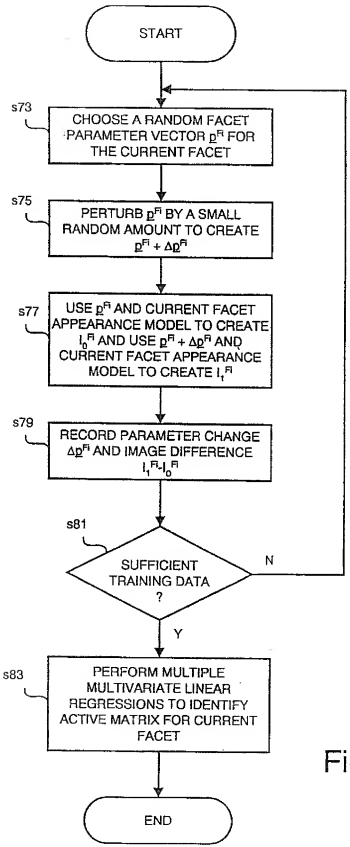

START
s73
CHOOSE A RANDOM FACET PARAMETER VECTOR $\underline{p}^{Fi}$ FOR THE CURRENT FACET
s75
PERTURB $\underline{p}^{Fi}$ BY A SMALL RANDOM AMOUNT TO CREATE $\underline{p}^{Fi} + \Delta\underline{p}^{Fi}$
s77
USE $\underline{p}^{Fi}$ AND CURRENT FACET APPEARANCE MODEL TO CREATE I_0^{Fi} AND USE $\underline{p}^{Fi} + \Delta\underline{p}^{Fi}$ AND CURRENT FACET APPEARANCE MODEL TO CREATE I_t^{Fi}
s79
RECORD PARAMETER CHANGE $\Delta\underline{p}^{Fi}$ AND IMAGE DIFFERENCE $I_t^{Fi} I_0^{Fi}$
s81
SUFFICIENT TRAINING DATA ?
N
Y
s83
PERFORM MULTIPLE MULTIVARIATE LINEAR REGRESSIONS TO IDENTIFY ACTIVE MATRIX FOR CURRENT FACET
END

Fig. 11a

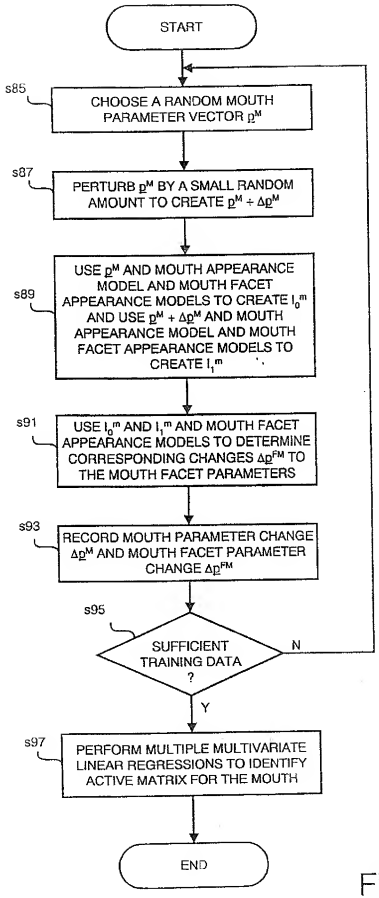

START
s85
CHOOSE A RANDOM MOUTH PARAMETER VECTOR $\underline{p}^M$
s87
PERTURB $\underline{p}^M$ BY A SMALL RANDOM AMOUNT TO CREATE $\underline{p}^M + \Delta\underline{p}^M$
s89
USE $\underline{p}^M$ AND MOUTH APPEARANCE MODEL AND MOUTH FACET APPEARANCE MODELS TO CREATE I_0^m AND USE $\underline{p}^M + \Delta\underline{p}^M$ AND MOUTH APPEARANCE MODEL AND MOUTH FACET APPEARANCE MODELS TO CREATE I_1^m
s91
USE I_0^m AND I_1^m AND MOUTH FACET APPEARANCE MODELS TO DETERMINE CORRESPONDING CHANGES $\Delta\underline{p}^{FM}$ TO THE MOUTH FACET PARAMETERS
s93
RECORD MOUTH PARAMETER CHANGE $\Delta\underline{p}^M$ AND MOUTH FACET PARAMETER CHANGE $\Delta\underline{p}^{FM}$
s95
SUFFICIENT TRAINING DATA ?
N
Y
s97
PERFORM MULTIPLE MULTIVARIATE LINEAR REGRESSIONS TO IDENTIFY ACTIVE MATRIX FOR THE MOUTH
END

Fig. 11b

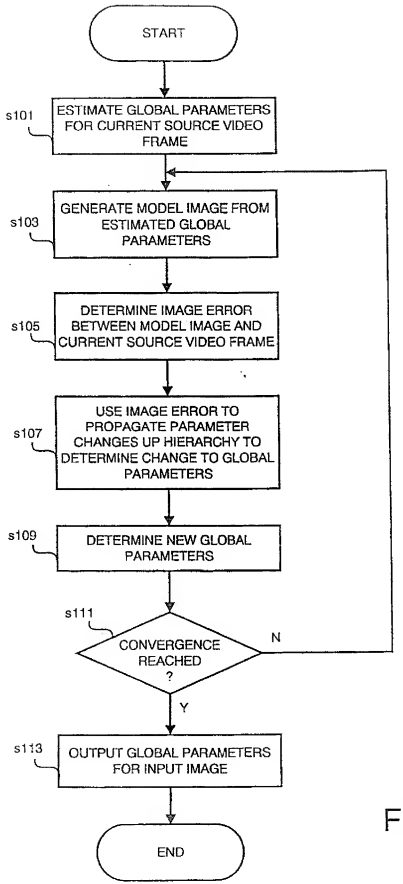

START
s101
ESTIMATE GLOBAL PARAMETERS FOR CURRENT SOURCE VIDEO FRAME
s103
GENERATE MODEL IMAGE FROM ESTIMATED GLOBAL PARAMETERS
s105
DETERMINE IMAGE ERROR BETWEEN MODEL IMAGE AND CURRENT SOURCE VIDEO FRAME
s107
USE IMAGE ERROR TO PROPAGATE PARAMETER CHANGES UP HIERARCHY TO DETERMINE CHANGE TO GLOBAL PARAMETERS
s109
DETERMINE NEW GLOBAL PARAMETERS
s111
CONVERGENCE REACHED ?
N
Y
s113
OUTPUT GLOBAL PARAMETERS FOR INPUT IMAGE
END

Fig. 12

Fig. 13a

Fig. 13b

Fig. 13c

Fig. 13d

Fig. 13e

INTERNATIONAL SEARCH REPORT

Internal Application No
PCT/GB 00/04411

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06T17/00 G06K9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06K G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC, IBM-TDB, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EDWARDS G J ET AL: "ADVANCES IN ACTIVE APPEARANCE MODELS" KERKYRA, GREECE, SEPT. 20 - 27, 1999,LOS ALMITOS, CA: IEEE COMP. PRESS,US, vol. CONF. 7, 1999, pages 137-142, XP000980072 ISBN: 0-7695-0165-6 abstract page 137, left-hand column, paragraph 1 -right-hand column, paragraph 2 --- -/--	1-32

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 April 2001	04/05/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk	

INTERNATIONAL SEARCH REPORT

Interna .al Application No
PCT/GB 00/04411

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	COOTES T F ET AL: "ACTIVE SHAPE MODELS-THEIR TRAINING AND APPLICATION" COMPUTER VISION AND IMAGE UNDERSTANDING,ACADEMIC PRESS,US, vol. 61, no. 1, January 1995 (1995-01), pages 38-59, XP000978654 ISSN: 1077-3142 abstract page 38, left-hand column, paragraph 1 -page 39, left-hand column, paragraph 1	1-32

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/37222 A1

(51) International Patent Classification⁷: G06T 17/00, G06K 9/00

(21) International Application Number: PCT/GB00/04411

(22) International Filing Date:
20 November 2000 (20.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9927314.6 18 November 1999 (18.11.1999) GB

(71) Applicant *(for all designated States except US)*: **ANTHROPICS TECHNOLOGY LIMITED** [GB/GB]; Ealing Studios, Ealing Green, London W5 5EP (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **NEWMAN, Rhys, Andrew** [GB/GB]; Anthropics Technology Limited, Ealing Studios, Ealing Green, London W5 5EP (GB). **WILES, Charles, Stephen** [GB/GB]; Anthropics Technology Limited, Ealing Studios, Ealing Green, London W5 5EP (GB). **WILLIAMS, Mark, Jonathan** [GB/GB]; Anthropics Technology Limited, Ealing Studios, Ealing Green, London W5 5EP (GB).

(74) Agents: **BERESFORD, Keith, Denis, Lewis** et al.; Beresford & Co, 2-5 Warwick Court, High Holborn, London WC1R 5DJ (GB).

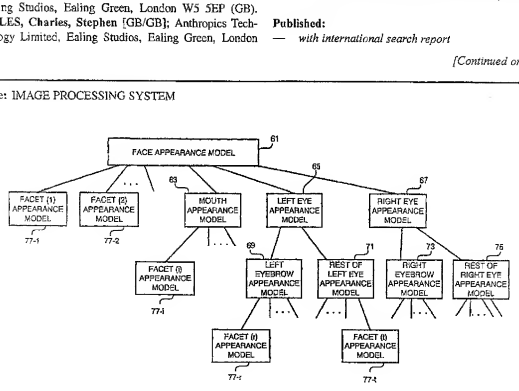
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

[Continued on next page]

(54) Title: IMAGE PROCESSING SYSTEM



(57) Abstract: A hierarchical parametric model is provided for modelling the appearance of objects, such as human faces. The hierarchical model can model both the shape and the texture of the object. The hierarchical model includes models for components of the object being modelled such that output parameters from one model are applied as input parameters to models which are lower in the hierarchy. This hierarchical model can be used, for example, for face tracking, for video compression, for 2D and 3D character

WO 01/37222 A1

(48) Date of publication of this corrected version:
9 August 2001

(15) Information about Correction:
see PCT Gazette No. 32/2001 of 9 August 2001, Section II

IMAGE PROCESSING SYSTEM

The present invention relates to the parametric modelling of the appearance of objects. The resulting model can be used, for example, to track the object, such as a human face, in a video sequence.

The use of parametric models for image interpretation and synthesis has become increasing popular. Cootes et al have shown in their paper entitled "Active Shape Models - Their Training and Application", Computer Vision and Image Understanding, Volume 61, No. 1, January, pages 38-59, 1995, how such parametric models can be used to model the variability of the shape and texture of human faces. They have mainly used these models for face recognition and tracking within video sequences, although they have also demonstrated that their model can be used to model the variability of other deformable objects, such as MRI scans of knee joints. The use of these models provides a basis for a broad range of applications since they explain the appearance of a given image in terms of a compact set of model parameters which can be used for higher levels of interpretation of the image. For example, when analysing face images, they can be used to characterise the identity, pose or expression of a face.

Using such models for image interpretation requires, however, a method of fitting them to new image data.

This involves identifying the model parameters that generate an image which best fits (according to some measure) the new input image. Typically this problem is one of minimising the sum of squares of pixel errors between the generated image and the input image. In their paper entitled "Estimating Coloured 3D Face Models from Single Images: An Example-Based Approach" Vetter and Blanz have proposed a stochastic gradient descent optimisation technique to identify the optimum model parameters for the new image. Although this technique can give very accurate results finding the locally optimal solution, they generally get stuck in local minima since the error surface for the problem of fitting an appearance model to an image is particularly rough containing many local minima. Therefore, this minimisation technique often fails to converge on the global minimum. An additional drawback of this technique is that it is very slow requiring several minutes to achieve convergence.

A faster, more robust technique known as the active appearance model was proposed by Edwards et al in the paper entitled "Interpreting Face Images using Active Appearance Models", published in the Third International Conference on Automatic Face and Gesture Recognition 1998, pages 300-305, Japan, April 1998. This technique uses a prior training stage in which the relationship between model parameter displacements and the resulting

change in image error is learnt. Although the method is much faster than direct optimisation techniques, it also requires fairly accurate initial model parameters if the search is to converge. Additionally, this technique does not guarantee that the optimum parameters will be found.

The appearance model proposed by Cootes et al includes a single appearance model matrix which linearly relates a set of parameters to corresponding image data. Blanz et al segmented the face into a number of completely independent appearance models, each of which is used to render a separate region of the face. The results are then merged using a general interpretation technique.

The present invention aims to provide an alternative way of modelling the appearance of objects which will allow subsequent image interpretation through appropriate processing of parameters generated for the image.

According to one aspect, the present invention provides a hierarchical parametric model for modelling the shape of an object, the model comprising data defining a hierarchical set of functions in which a function in a top layer of the hierarchy is operable to generate a set of output parameters from a set of input parameters and in which one or more functions in a bottom layer of the hierarchy are operable to receive parameters output from one or more functions from a higher layer of the

hierarchy and to generate therefrom the relative positions of a plurality of predetermined points on the object. Such a hierarchical parametric model has the advantage that small changes in some parts of the object can still be modelled by the parameters, even though they are significantly smaller than variations in other less important parts of the object. This model can be used for face tracking, video compression, 2D and 3D character generation, face recognition for security purposes, image editing etc.

According to another aspect, the present invention provides an apparatus and method of determining a set of appearance parameters representative of the appearance of an object, the method comprising the steps of storing a hierarchical parametric model such as the one discussed above and at least one function which relates a change in input parameters to an error between actual appearance data for the object and appearance data determined from the set of input parameters and the parametric model; initially receiving a current set of input parameters for the object; determining appearance data for the object from the current set of input parameters and the stored parametric model; determining the error between the actual appearance data of the object and the appearance data determined from the current set of input parameters; determining a change in the input parameters using the at least one stored function and said determined error; and

updating the current set of input parameters with the determined change in the input parameters.

An exemplary embodiment of the present invention will now be described with reference to the accompanying drawings in which:

Figure 1 is a schematic block diagram illustrating a general arrangement of a computer system which can be programmed to implement the present invention;

Figure 2 is a block diagram of an appearance model generation unit which receives some of the image frames of a source video sequence together with a target image frame and generates therefrom an appearance model;

Figure 3 is a block diagram of a target video sequence generation unit which generates a target video sequence from a source video sequence using a set of stored difference parameters;

Figure 4 is a flow chart illustrating the processing steps which the target video sequence generation unit shown in Figure 3 performs to generate the target video sequence;

Figure 5 schematically illustrates the form of a hierarchical appearance model generated in one embodiment

of the invention;

Figure 6 shows a head with a mesh of triangular facets placed over the head and whose positions are defined by the position of landmark points at the corners of the facets;

Figure 7 is a flow chart illustrating the processing steps required to generate a facet appearance model from the training images;

Figure 8 schematically illustrates the way in which a transformation is defined between a facet in a training image and a predefined shape of facet which allows texture information to be extracted from the facet;

Figure 9 is a flow chart illustrating the main processing steps involved in determining an appearance model for the mouth using the appearance models for the facets which appear in the mouth and using the training images;

Figure 10 schematically illustrates the way in which training images are used to determine some of the appearance models which form the hierarchical appearance model illustrated in Figure 5;

Figure 11a is a flow chart illustrating the processing steps performed during a training routine to identify an

Active matrix associated with a current facet;

Figure 11b is a flow chart illustrating the processing steps performed during a training routine to identify an Active matrix associated with the mouth;

Figure 12 is a flow chart illustrating the processing steps involved in determining a set of parameters which define the appearance of a face within a input image;

Figure 13a shows three frames of an example source video sequence which is applied to the target video sequence generation unit shown in Figure 4;

Figure 13b shows an example target image used to generate a set of difference parameters used by the target video sequence generation unit shown in Figure 4;

Figure 13c shows a corresponding three frames from a target video sequence generated by the target video sequence generation unit shown in Figure 4 from the three frames of the source video sequence shown in Figure 13a using the difference parameters generated using the target image shown in Figure 13b;

Figure 13d shows a second example of a target image used to generate a set of difference parameters for use by the target video sequence generation unit shown in Figure 4;

and

Figure 13e shows the corresponding three frames from the target video sequence generated by the target video sequence generation unit shown in Figure 4 when the three frames of the source video sequence shown in Figure 13a are input to the target video sequence generation unit together with the difference parameters calculated using the target image shown in Figure 13d.

Figure 1 is an image processing apparatus according to an embodiment of the present invention. The apparatus comprises a computer 1 having a central processing unit (CPU) 3 connected to a memory 5 which is operable to store a program defining the sequence of operations of the CPU 3 and to store object and image data used in calculations by the CPU 3. Coupled to an input port of the CPU 3 there is an input device 7, which in this embodiment comprises a keyboard and a computer mouse. Instead of, or in addition to the computer mouse, another position sensitive input device (pointing device) such as a digitiser with associated stylus may be used.

A frame buffer 9 is also provided and is coupled to the CPU 3 and comprises a memory unit (not shown) arranged to store image data relating to at least one image, for example by providing one (or several) memory location(s) per pixel of the image. The value stored in the frame

buffer for each pixel defines the colour or intensity of that pixel in the image. In this embodiment, the images are represented by 2-D arrays of pixels, and are conveniently described in terms of Cartesian coordinates, so that the position of a given pixel can be described by a pair of x-y coordinates. This representation is convenient since the image is displayed on a raster scan display 11. Therefore, the x-coordinate maps to the distance along the line of the display and the y-coordinate maps to the number of the line. The frame buffer 9 has sufficient memory capacity to store at least one image. For example, for an image having a resolution of 1000 x 1000 pixels, the frame buffer 9 includes 10^6 pixel locations, each addressable directly or indirectly in terms of a pixel coordinate x,y.

In this embodiment, a video tape recorder (VTR) 13 is also coupled to the frame buffer 9, for recording the image or sequence of images displayed on the display 11. A mass storage device 15, such as a hard disc drive, having a high data storage capacity is also provided and coupled to the memory 5. Also coupled to the memory 5 is a floppy disc drive 17 which is operable to accept removable data storage media, such as a floppy disc 19 and to transfer data stored thereon to the memory 5. The memory 5 is also coupled to a printer 21 so that generated images can be output in paper form, an image input device 23 such as a scanner or video camera and a

modem 25 so that input images and output images can be received from and transmitted to remote computer terminals via a data network, such as the Internet. The CPU 3, memory 5, frame buffer 9, display unit 11 and mass storage device 13 may be commercially available as a complete system, for example as an IBM compatible personal computer (PC) or a workstation such as the Sparc station available from Sun Microsystems.

A number of embodiments of the invention can be supplied commercially in the form of programs stored on a floppy disc 19 or on other mediums, or as signals transmitted over a data link, such as the Internet, so that the receiving hardware becomes reconfigured into an apparatus embodying the present invention.

In this embodiment, the computer 1 is programmed to receive a source video sequence input by the image input device 23 and to generate a target video sequence from the source video sequence using a target image. In this embodiment, the source video sequence is a video clip of an actor acting out a scene, the target image is an image of a second actor and the resulting target video sequence is a video sequence showing the second actor acting out the scene. The way in which this is achieved will now be briefly described with reference to Figures 2 to 4.

In this embodiment, in order to generate the target video

sequence from the source video sequence, a hierarchical parametric appearance model which models the variability of shape and texture of the head images is used. This appearance model makes use of the fact that some prior knowledge is available about the contents of head images in order to facilitate their modelling. For example, it can be assumed that two frontal images of a human face will each include eyes, a nose and a mouth. In this embodiment, as shown in Figure 2, the hierarchical parametric appearance model 35 is generated by an appearance model generation unit 31 from training images which are stored in an image database 32. In this embodiment, all the training images are colour images having 500 x 500 pixels, with each pixel having a red, green and a blue pixel value. The resulting appearance model 35 is a parameterisation of the appearance of the class of head images defined by the heads in the training images, so that a relatively small number of parameters (for example 50) can describe the detailed (pixel level) appearance of a head image from the class. In particular, the hierarchical appearance model 35 defines a function (F) such that:

$$I = F(\underline{p}) \qquad (1)$$

where $\underline{p}$ is the set of appearance parameters (written in vector notation) which generates, through the hierarchical appearance model (F), the face image I. The

structure of the hierarchical appearance model used in this embodiment will be described later.

Once the hierarchical appearance model 35 has been determined, a target video sequence can be generated from a source video sequence. As shown in Figure 3, the source video sequence is input to a target video sequence generation unit 51 which processes the source video sequence using a set of difference parameters 53 to generate and to output the target video sequence. The difference parameters 53 are determined by subtracting the appearance parameters which are generated for the first actor's head in one of the source video frames, from the appearance parameters which are generated for the second actor's head in the target image. The way in which these appearance parameters are determined for these images will be described later. In order that these difference parameters only represent differences in the general shape and colour texture of the two actors' heads, the pose and facial expression of the first actor's head in the source video frame used should match, as closely as possible, the pose and facial expression of the second actor's head in the target image.

The processing steps required to generate the target video sequence from the source video sequence will now be described in more detail with reference to Figure 4. As shown, in step s1, the appearance parameters (p_a^i) for

the first actor's head in the current video frame (I_s^i) are automatically calculated. The way that this is achieved will be described later. Then, in step s3, the difference parameters ($\underline{p}_{dif}$) are added to the appearance parameters for the first actor's head in the current video frame to generate:

$$\underline{p}_{mod}^{i} = \underline{p}_{s}^{i} + \underline{p}_{dif} \qquad (2)$$

The resulting appearance parameters ($\underline{p}_{mod}^{i}$) are then used, in step s5, to regenerate the head for the current target video frame. In particular, the modified appearance parameters are inserted into equation (1) above to regenerate a modified head image which is then composited, in step s7, into the source video frame to generate the corresponding target video frame. A check is then made, in step s9, to determine whether or not there are any more source video frames. If there are, then the processing returns to step s1 where the procedure described above is repeated for the next source video frame. If there are no more source video frames, then the processing ends.

Figure 13 illustrates the results of this animation technique (although showing black and white images and not colour). In particular, Figure 13a shows three frames of the source video sequence, Figure 13b shows the target image (which in this embodiment is computer

generated) and Figure 13c shows the corresponding three frames of the target video sequence obtained in the manner described above. As can be seen, an animated sequence of the computer generated character has been generated from a video clip of a real person and a single image of the computer generated character.

HIERARCHICAL APPEARANCE MODEL

In the systems described by Cootes et al and Blanz et al, the parametric model is created by placing a number of landmark points on a training image and then identifying the same landmark points on the other training images in order to identify how the location of and the pixel values around the landmark points vary within the training images. A principal component analysis is then performed on the matrix which consists of vectors of the landmark points. This PCA yields a set of Eigenvectors which describe the directions of greatest variation along which the landmark points change. Their appearance model includes the linear combination of the Eigenvectors plus parameters for translation, rotation and scaling. This single appearance model relates a compact set of appearance parameters to pixel values.

In this embodiment, rather than having a single appearance model for the object, a hierarchical appearance model comprising several appearance models which model variations in components of the object is

used. For example, in the case of human faces, the hierarchical appearance model may include an appearance model for the mouth, one for the left eye, one for the right eye and one for the nose. Since it may be possible to model various components of the object, the particular hierarchical structure which will be used for a particular object and application must first of all be defined by the system designer.

Figure 5 schematically illustrates the structure of the hierarchical appearance model used in this embodiment. As shown, at the top of the hierarchy there is a general face appearance model 61. Beneath the face appearance model there is a mouth appearance model 63, a left eye appearance model 65, a right eye appearance model 67, a left eyebrow appearance model 69, a rest of left eye appearance model 71, a right eyebrow appearance model 73, a rest of right eye appearance model 75 and, in this embodiment, a facet appearance model for each facet defined in the training images. Figure 6 shows the head of a training image in which the set of landmark points has been placed at the appropriate points on the head. As shown, in this embodiment, there are one hundred and forty-eight triangular areas or facets defined by the positions of the landmark points. Therefore, in this embodiment, there are one hundred and forty-eight facet appearance models 77.

The face appearance model 61 operates to relate a small number of "global" appearance parameters to a further set of appearance parameters, some of which are input to facet appearance models 77, some of which are input to the mouth appearance model 63, some of which are input to the left eye appearance model 65 and the rest of which are input to the right eye appearance model 67. The facet appearance models 77 operate to relate the input parameters received from the appearance model which is above it in the hierarchy into corresponding pixel values for that facet. The mouth appearance model 63 is operable to relate the parameters it receives from the face appearance model 61 into a further set of appearance parameters, respective ones of which are output to the respective facet appearance models 77 for the facets which are associated with the mouth. Similarly, the left and right eye appearance models 65 and 67 operate to relate the parameters it receives from the face appearance model 61 into a further set of appearance parameters, some of which are input to the appropriate eyebrow appearance model and the rest of which are input to the appropriate rest of eye appearance model. These appearance models in turn convert these parameters into parameters for input to the facet appearance models associated with the facets which appear in the left and right eyes respectively. In this way, a small compact set of "global" appearance parameters input to the face appearance model 61 can filter through the hierarchical

structure illustrated in Figure 5 to generate a set of pixel values for all the facets in a head which can then be used to regenerate the image of the head.

The way in which the individual appearance models of this hierarchical appearance model are generated in this embodiment will now be described with reference to Figures 6 to 10.

In this embodiment, each of the training images stored in the image database 32 is labelled with eighty six landmark points. In this embodiment, this is performed manually by the user via the user interface 33. In particular, each training image is displayed on the display 11 and the user places the landmark points over the head in the training image. These points delineate the main features in the head, such as the position of the hairline, neck, eyes, nose, ears and mouth. In order to compare training faces, each landmark point is associated with the same point on each face. In this embodiment, the following landmark points are used:

Landmark Point	Associated Position	Landmark Point	Associated Position
LP_1	Left corner of left eye	LP_{44}	Eye, bottom
LP_2	Right corner of right eye	LP_{45}	Eye, top
LP_3	Chin, bottom	LP_{46}	Eye, bottom
LP_4	Right corner of left eye	LP_{47}	Eyebrow, lower

Landmark Point	Associated Position	Landmark Point	Associated Position
LP_5	Left corner of right eye	LP_{48}	Eyebrow, upper
LP_6	Mouth, left	LP_{49}	Cheek, left
LP_7	Mouth, right	LP_{50}	Cheek, right
LP_8	Nose, bottom	LP_{51}	Eyebrow, lower
LP_9	Nose, between eyes	LP_{52}	Eyebrow, upper
LP_{10}	Upper lip, top	LP_{53}	Eyebrow, lower
LP_{11}	Lower lip, bottom	LP_{54}	Eyebrow, upper
LP_{12}	Neck, left, top	LP_{55}	Eyebrow, lower
LP_{13}	Neck, right, top	LP_{56}	Eyebrow, upper
LP_{14}	Face edge left, level with nose	LP_{57}	Eyebrow, lower
LP_{15}	Face edge	LP_{58}	Eyebrow, upper
LP_{16}	Face edge right, level with nose	LP_{59}	Eyebrow, lower
LP_{17}	Face edge	LP_{60}	Eyebrow, upper
LP_{18}	Top of head	LP_{61}	Eyebrow, lower
LP_{19}	Hair edge	LP_{62}	Lower lip, top
LP_{20}	Hair edge	LP_{63}	Centre forehead
LP_{21}	Hair edge	LP_{64}	Upper lip, top left
LP_{22}	Hair edge	LP_{65}	Upper lip, top right
LP_{23}	Hair edge	LP_{66}	Lower lip, bottom right
LP_{24}	Hair edge	LP_{67}	Lower lip, bottom left
LP_{25}	Hair edge	LP_{68}	Eye, top left
LP_{26}	Hair edge	LP_{69}	Eye, top right
LP_{27}	Hair edge	LP_{70}	Eye, bottom right
LP_{28}	Hair edge	LP_{71}	Eye, bottom left
LP_{29}	Bottom, far left	LP_{72}	Eye, top left
LP_{30}	Bottom, far right	LP_{73}	Eye, top right
LP_{31}	Shoulder	LP_{74}	Eye, bottom right

Landmark Point	Associated Position	Landmark Point	Associated Position
LP_{32}	Shoulder	LP_{75}	Eye, bottom left
LP_{33}	Bottom, left	LP_{76}	Lower lip, top left
LP_{34}	Bottom, middle	LP_{77}	Lower lip, top right
LP_{35}	Bottom, right	LP_{78}	Chin, left
LP_{36}	Left forehead	LP_{79}	Chin, right
LP_{37}	Right forehead	LP_{80}	Neck, left
LP_{38}	Centre, between eyebrows	LP_{81}	Neckline, left
LP_{39}	Nose, left	LP_{82}	Neckline
LP_{40}	Nose, right	LP_{83}	Neckline, right
LP_{41}	Nose edge, left	LP_{84}	Neck, right
LP_{42}	Nose edge, right	LP_{85}	Hair edge
LP_{43}	Eye, top	LP_{86}	Hair edge

The result of the manual placement of the landmark points is a table of landmark points for each training image, which identifies the (x, y) coordinate of each landmark point within the image. As shown in Figure 6, these landmark points are also used to define the location of predetermined triangular facets or areas within the training image.

FACET APPEARANCE MODEL

Figure 7 shows a flow chart illustrating the main processing steps involved in this embodiment in determining a facet appearance model for facet (i). As shown, in step s61, the system determines, for each training image, the apex coordinates of facet (i) and

texture values from within facet (i). In order to sample texture from within the facet at corresponding points within each training facet, a transformation which transforms the facet onto a reference facet is determined. Figure 8 illustrates this transformation. In particular, Figure 8 shows facet f_i^v taken from the V-th training image, which is defined by the landmark points (x_1^v, y_1^v), (x_2^v, y_2^v) and (x_3^v, y_3^v). The transformation (T_i^v) which transforms those coordinates onto coordinates (0,0), (1,0) and (0,1) is determined. In this embodiment, the texture information extracted from each training facet is defined by the regular array of pixels shown in the reference facet. In order to determine the corresponding red, green and blue pixel values in the training image, the inverse transformation $([T_i^v]^{-1})$ is used to transform the pixel locations in the reference facet, into corresponding locations in the training facet, from which the RGB pixel values are determined. In this embodiment, this transformation may not result in an exact correspondence with a single image pixel location since the pixel resolution in the actual facet may be different to the resolution in the reference facet. In this embodiment, the texture information (RGB pixel values) which is determined is obtained by interpolating between the surrounding image RGB pixel values. In this embodiment, there are fifty pixels in the regular array of pixels in the reference facet. Therefore, fifty RGB pixel values are extracted for each

training facet. The texture information for facet (i) from the V-th training image can then be represented by a vector (t^{iv}) of the form:

$$\underline{t}^{iv} = [t_1^{iv}, t_2^{iv}, t_3^{iv} \ \ldots \ t_{50}^{iv}]^T$$

where t_1^{iv} is the RGB texture information for the first reference pixel extracted from facet (i) in the V-th training image etc.

In this embodiment, the facet appearance models 77 treat shape and texture separately. Therefore, in step s63, the system performs a principal component analysis (PCA) on the set of texture training vectors generated in step s61. For a more detailed discussion of principal component analysis, the reader is referred to the book by W. J. Krzanowski entitled "Principles of Multivariate Analysis - A User's Perspective" 1998, Oxford Statistical Science Series. As those skilled in the art will appreciate, this principal component analysis determines all possible modes of variation within the training texture vectors. However, since each of the facets is associated with a similar point on the face, most of the variation within the data can be explained by a few modes of variation. The result of the principal component analysis is a facet texture appearance model (defined by matrix F_1) which relates a vector of facet texture parameters to a vector of texture pixel values, by:

$$\underline{p}_v^{fit} = F_i(\underline{t}^{iv} - \overline{\underline{t}}^i) \quad (3)$$

where $\underline{t}^{iv}$ is the RGB texture vector defined above, $\overline{\underline{t}}^i$ is the mean RGB texture vector for facet (i), F_i is a matrix which defines the facet texture appearance model for facet (i) and $\underline{p}^{Fit}{}_v$ is a vector of the facet texture parameters which describes the RGB texture vector $\underline{t}^{iv}$. The matrix F_i describes the main modes of variation of the texture within the training facets; and the vector of facet texture parameters ($\underline{p}^{Fit}{}_v$) for a given input facet has a parameter associated with each mode of variation whose value relates the texture of the input facet to the corresponding mode of variation.

As those skilled in the art will appreciate, for facets which describe fairly constant parts of the face, such as the chin or cheeks, very few parameters will be needed to model the variability within the training images. However, facets which are associated with areas of the face where there is a large amount of variability (such as facets which form part of the eye), will require a larger number of facet texture parameters to describe the variability within the training images. Therefore, in step s65, the system determines how many texture parameters are needed for the current facet and stores the appropriate facet appearance model matrix.

In addition to being able to determine a set of texture

parameters $\underline{p}^{Fit}_{v}$ for a given texture vector $\underline{t}^{iv}$, equation (3) can be solved with respect to the texture vector $\underline{t}^{iv}$ to give:

$$\underline{t}^{iv} = \overline{\underline{t}}^{i} - F_i^T \underline{p}_v^{Fit} \qquad (4)$$

since $F_i F_i^T$ equals the identity matrix. Therefore, by modifying the set of texture parameters ($\underline{p}^{Fit}$) within suitable limits, new textures for facet (i) can be generated which are similar to those in the training set.

Once the above procedure has been performed for each of the one hundred and forty-eight facets in the training images, a facet texture appearance model will have been generated for each of those facets. In this embodiment, the facet appearance model does not compress the parameters defining the shape of the facets, since only six parameters are needed to define the shape of each facet - two parameters for each (x,y) coordinate of the facet's apexes.

MOUTH APPEARANCE MODEL

Figure 9 shows a flow chart illustrating the main processing steps required in order to generate the mouth appearance model 63. As shown, in step s67, the system uses the facet appearance models for the facets which form part of the mouth to generate shape and texture parameters from those facets for each training image.

Therefore, referring to Figure 10, the mouth appearance model 63 will receive texture and shape parameters from the facet appearance model for facet (i), facet (j) and facet (n) for the corresponding facets in each of the training images 79. As illustrated in Figure 10, the appearance model for facet (i) is operable to generate, for each training image, six shape parameters (corresponding to the three (x,y) coordinates of the apexes of facet (i)) and six texture parameters. Similarly, the appearance model for facet (j) is operable to generate, for each training image, six shape parameters and four texture parameters and the appearance model for facet (n) is operable to generate, for each training image, six shape parameters and three texture parameters.

The processing then proceeds to step s69 where the system performs a principal component analysis on the shape and texture parameters generated for the training images by the facet appearance models associated with the mouth. In this embodiment, the mouth appearance model 63 treats the shape and texture separately. In particular, for each training image, the system concatenates the six shape parameters for the facets associated with the mouth to form the following shape vector:

$$\underline{p}^{FMs} = [x_1^{fi}, y_1^{fi}, x_2^{fi}, y_2^{fi}, x_3^{fi}, y_3^{fi} \; : \; x_1^{fj}, y_1^{fj}, x_2^{fj}, y_2^{fj} \; \ldots]^T$$

and concatenates the facet texture parameters output by the facet appearance models associated with the mouth to form the following texture vector:

$$\underline{P}^{FMt} = [p_1^{Fit}, p_2^{Fit} \ldots p_6^{Fit} : p_1^{Fjt}, p_2^{Fjt} \ldots p_4^{Fjt} : p_1^{Fkt}, \ldots]^T$$

The system then performs a principal component analysis on the shape vectors generated by all the training images to generate a shape appearance model for the mouth (defined by matrix M_s) which relates each mouth shape vector to a corresponding vector of shape mouth parameters by:

$$\underline{P}_v^{Ms} = M_s(\underline{p}_v^{FMs} - \overline{\underline{P}}^{FMs}) \qquad (5)$$

where $\underline{P}^{FMs}{}_v$ is the mouth shape vector for the mouth in the V-th training image, $\overline{\underline{P}}^{FMs}$ is the mean mouth shape vector from the training vectors and $\underline{P}^{Ms}{}_v$ is a vector of mouth shape parameters for the mouth shape vector $\underline{P}^{FMs}{}_v$. The mouth shape model, defined by matrix M_s, describes the main modes of variation of the shape of the mouths within the training images; and the vector of mouth shape parameters ($\underline{P}^{Ms}{}_v$) for the mouth in the V-th training image has a parameter associated with each mode of variation whose value relates the shape of the input mouth to the corresponding mode of variation.

As with the facet appearance models, equation (5) above

can be rewritten with respect to the mouth shape vector $\underline{p}^{FMs}{}_v$ to give:

$$p_v^{FMs} = \overline{p}^{FMs} - M_s^T \underline{p}_v^{Ms} \qquad (6)$$

since M_s $M_s{}^T$ equals the identity matrix. Therefore, by modifying the mouth shape parameters, new mouth shapes can be generated which will be similar to those in the training set.

The system then performs a principal component analysis on the mouth texture parameter vectors ($\underline{p}^{FMt}$) which are generated for the training images. This principal component analysis generates a mouth texture model (defined by matrix M_t) which relates each of the facet texture parameter vectors for the facets associated with the mouth, to a corresponding vector of mouth texture parameters, by:

$$\underline{p}_v^{Mt} = M_t(\underline{p}_v^{FMt} - \overline{p}^{FMt}) \qquad (7)$$

where $\underline{p}^{FMt}{}_v$ is a vector of mouth facet texture parameters generated by the facet appearance models associated with the mouth for the mouth in the V-th training image; $\overline{p}^{FMt}$ is the mean vector of mouth facet texture parameters from the training vectors and $\underline{p}^{Mt}{}_v$ is a vector of mouth texture parameters for the facet texture parameters $\underline{p}^{FMt}{}_v$. The matrix M_t describes the main modes of variation within

the training images of the facet texture parameters generated by the facet appearance models which are associated with the mouth; and the vector of mouth texture parameters (p^{mt}_v) has a parameter associated with each of those modes of variation whose value relates the texture of the input mouth to the corresponding mode of variation.

The processing then proceeds to step s71 shown in Figure 9 where the system determines the number of shape parameters and texture parameters needed to describe the training data received from the facet appearance models which are associated with the mouth. As shown in Figure 10, in this embodiment, the mouth appearance model 63 requires five shape parameters and four texture parameters to be able to model most of this variation. The system therefore stores the appropriate mouth shape and texture appearance model matrices for subsequent use.

As those skilled in the art will appreciate, a similar procedure is performed to determine each of the appearance models shown in Figure 5, starting from the facet appearance models at the base of the hierarchy. A further description of how these remaining appearance models are determined will, therefore, not be given here. The resulting hierarchical appearance model allows a small number of global face appearance parameters to be input to the face appearance model 61, which generates

further parameters which propagate down through the hierarchical model structure until facet pixel values are generated, from which an image which corresponds to the global appearance parameters can be generated.

AUTOMATIC GENERATION OF APPEARANCE PARAMETERS

In the description given above of the way in which the appearance models are generated, appearance parameters for an image were generated from a manual placement of a number of landmark points over the image. However, during use of the appearance model to track the first actor's head in the source video sequence and during the calculation of the difference parameters (p_{diff}), the appearance parameters for the heads in the input images were automatically calculated. This task involves finding the set of global appearance parameters $\underline{p}$ which best describe the pixels in view. This problem is complicated because the inverse of each of the appearance models in the hierarchical appearance model is not necessarily one-to-one. In this embodiment, the appearance parameters for the head in an input image are calculated in a two-step process. In the first step, an initial set of global appearance parameters for the head in the current frame (I_s^i) is found using a simple and rapid technique. For all but the first frame of the source video sequence, this is achieved by simply using the appearance parameters from the preceding video frame (I_s^{i-1}) before modification in step s3 (i.e. parameters

$\underline{p}_s^{i-1}$). In this embodiment, the global appearance parameters ($\underline{p}$) effectively define the shape and colour texture of the head. For the first frame and for the target image the initial estimate of the appearance parameters is set to the mean set of appearance parameters and the scale, position and orientation is initially estimated by the user manually placing the mean head over the head in the image.

In the second step, an iterative technique is used in order to make fine adjustments to the initial estimate of the appearance parameters. The adjustments are made in an attempt to minimise the difference between the head described by the global appearance parameters (the model head) and the head in the current video frame (the image head). With 50 appearance parameters, this represents a difficult optimisation problem. This can be performed by using a standard steepest descent optimisation technique to iteratively reduce the mean squared error between the given image pixels and those predicted by a particular set of appearance parameter values. In particular, minimising the following error function $E(\underline{p})$:

$$E(\underline{p}) = [I^a - F(\underline{p})]^T [I^a - F(\underline{p})] \quad (8)$$

where I^a is a vector of actual image RGB pixel values at the locations where the appearance model predicts values

(the appearance model does not predict all pixel values since it ignores background pixels and only predicts a subsample of pixel values within the object being modelled) and $F(\underline{p})$ is the vector of image RGB pixel values predicted by the hierarchical appearance model. As those skilled in the art will appreciate, $E(\underline{p})$ will only be zero when the model head (i.e. $F(\underline{p})$) predicts the actual image head (I^a) exactly. Standard steepest descent optimisation techniques stipulate that a step in the direction $-\nabla E(\underline{p})$ should result in a reduction in the error function $E(\underline{p})$, provided the error function is well behaved. Therefore, the change ($\Delta\underline{p}$) in the set of parameter values should be:

$$\Delta\underline{p} = 2[\nabla F(\underline{p})]^T[I^a - F(\underline{p})] \quad (9)$$

which requires the calculation of the differential of the appearance model, i.e. $\nabla F(\underline{p})$.

The technique described by Edwards et al assumes that, on average over the whole parameter space, $\nabla F(\underline{p})$ is constant. The update equation then becomes:

$$\Delta\underline{p} = A[I^a - F(\underline{p})] \quad (10)$$

for some constant matrix A (referred to as the "Active matrix") which is determined beforehand during a training routine. In this embodiment, rather than using a single

constant matrix associated with the entire hierarchical appearance model, an Active matrix is determined and used for each of the individual appearance models which form part of the hierarchical appearance model. The way in which these Active matrices are determined in this embodiment will now be described with reference to Figures 11a and 11b, which illustrate the processing steps performed to generate the Active matrix for each facet appearance model and the Active matrix for the mouth appearance model.

As shown in Figure 11a, in step s73, the system chooses a random facet parameter vector ($\underline{p}^{Fi}$) for the current facet (i) and then, in s75, perturbs this facet parameter vector by a small random amount to create $\underline{p}^{Fi} + \Delta\underline{p}^{Fi}$. In this embodiment, the facet parameter vectors include not only the texture parameters, but also the six shape parameters which define the (x,y) coordinates of the facet's location within the image. The processing then proceeds to step s77 where the system uses the parameter vector $\underline{p}^{Fi}$ and the perturbed parameter vector $\underline{p}^{Fi} + \Delta\underline{p}^{Fi}$ to create model images I_0^{Fi} and I_1^{Fi} respectively. The processing then proceeds to step s79 where the system records the parameter change $\Delta\underline{p}^{Fi}$ and image difference $I_1^{Fi} - I_0^{Fi}$. Then in step s81, the system determines whether or not there is sufficient training data for the current facet. If there is not then the processing returns to step s21. Once sufficient training data has been

generated, the processing proceeds to step s83 where the system performs multiple multivariate linear regressions on the data for the current facet to identify an Active matrix (A_{Fi}) for the current facet.

Figure 11b shows the processing steps required to calculate the Active matrix for the mouth appearance model. As shown, in step s85, the system chooses a random mouth parameter vector $\underline{p}^M$. In this embodiment, this vector includes both the mouth shape parameters and the mouth texture parameters. Then, in step s87, the system perturbs this mouth parameter vector by a small random amount to create $\underline{p}^M + \Delta\underline{p}^M$. The processing then proceeds to step s89 where the system uses the mouth parameter vectors $\underline{p}^M$ and the perturbed mouth parameter vector $\underline{p}^M + \Delta\underline{p}^M$ to create model images I_0^m and I_1^m respectively, using the mouth appearance model and the facet appearance models associated with the mouth. The processing then proceeds to step s91 where the facet appearance models associated with the mouth are used again to transform the mouth model images I_0^m and I_1^m into corresponding facet appearance parameters $\underline{p}_0^{FM}$ and $\underline{p}_1^{FM}$, which are then subtracted to determine the corresponding change $\Delta\underline{p}^{FM}$ in the mouth facet parameters. The processing then proceeds to step s93 where the system records the mouth parameter change $\Delta\underline{p}^M$ and the mouth facet parameter change $\Delta\underline{p}^{FM}$. The processing then proceeds to step s95 where the system determines whether

or not there is sufficient training data. If there is not, then the processing returns to step s85. Once sufficient training data has been generated, the processing proceeds to step s97, where the system performs multiple multivariate linear regressions on the training data for the mouth to identify the Active matrix (A_M) for the mouth which relates changes in mouth parameters $\Delta\underline{p}^M$ to changes in facet parameters $\Delta\underline{p}^{FM}$ for the facets associated with the mouth.

As those skilled in the art will appreciate, a similar processing technique is used in order to identify the Active matrix for each of the appearance models shown in Figure 5.

Once the Active matrices have been determined for the hierarchical appearance model, they can then be used to iteratively update a current estimate of a set of appearance parameters for an input image. Figure 12 illustrates the processing steps performed in this iterative routine for the current source video frame. As shown, in step s101, the system initially estimates a set of global parameters for the head in the current source video frame. The processing then proceeds to step s103 where the system generates a model image from the estimated global parameters and the hierarchical appearance model. The system then proceeds to step s105 where it determines the image error between the model

image and the current source video frame. Then, in step s107, the system uses this image error to propagate parameter changes up the hierarchy of the hierarchical appearance model using the stored Active matrices to determine a change in the global parameters. This change in global parameters is then used, in step s109, to update the current global parameters for the current source video frame. The system then determines, in step s111, whether or not convergence has been reached by comparing the error obtained from equation (8) using the updated global parameters with a predetermined threshold (Th). If convergence has not been reached, then the processing returns to step s103. Once convergence is reached, the processing proceeds to step s113, where the current global appearance parameters are output as the global appearance parameters for the current source video frame and then the processing ends.

ALTERNATIVE EMBODIMENTS

In the above embodiment, the same hierarchical model structure was used to model the variation in the shape and texture within the training images. As those skilled in the art will appreciate, one model hierarchy can be used to model the shape variation and a different model hierarchy can be used to model the texture variation. Alternatively still, rather than separating the shape and texture parameters, each of the appearance models within the hierarchical model may model the combined variation

of the shape and texture within the training images.

In the above embodiments, a facet appearance model was generated for each facet defined within the training images. As those skilled in the art will appreciate, many of the facets may be grouped together such that a single facet appearance model is generated for those facets. In one form of such an embodiment, a single facet appearance model may be determined which models the variability of texture within each facet of the training images.

In the above embodiments, the same amount of texture information was extracted from each facet within the training images. In particular, fifty RGB texture values were extracted from each training facet. In an alternative embodiment, the amount of texture information extracted from each facet may vary in dependence upon the size of the facet. For example, more texture information may be extracted from larger facets or more texture information may be extracted from facets associated with important features of the face, such as the mouth, eyes or nose.

In the above embodiments, each appearance model was determined from a principal component analysis of a set of training data. This principal component analysis determines a linear relationship between the training

data and a set of model parameters. As those skilled in the art will appreciate, techniques other than principal component analysis can be used to determine a parametric model which relates a set of parameters to the training data. This model may define a non-linear relationship between the training data and the model parameters. For example, one or more of the models within the hierarchy may comprise a neural network which relates the set of input parameters to the training data.

In the above embodiments, a principal component analysis was performed on a set of training data in order to identify a relatively small number of parameters which describe the main modes of variation within the training data. This allows a relatively small number of input parameters to be able to generate a larger set of output parameters from the model. However, as those skilled in the art will appreciate, this is not essential. One or more of the appearance models may act as transformation models in which the number input parameters is the same as or greater than the number of output parameters. This can be used to generate a set of input parameters which can be changed by the user in some intuitive way. For example, in order to identify parameters which have a linear relationship with features in the object, such as a parameter that linearly changes the amount of smile within a face image.

In the above embodiments, a set of Active matrices were used in order to identify automatically a set of appearance parameters for an input image. As those skilled in the art will appreciate, rather than having separate Active matrices for each of the components in the hierarchical appearance model, a global Active matrix may be used instead. Further, although both the shape and grey level parameters were used in order to derive the Active matrices, suitable Active matrices can be determined using just the shape information.

In the above embodiments, the variation in both the shape and texture within the training images were modelled. As those skilled in the art will appreciate, this hierarchical modelling technique can be used to model only the shape of the objects within the training images. Such a shape model could be then used to track objects within a video sequence.

In the first embodiment, the target image illustrated a computer generated head. This is not essential. For example, the target image might be a hand-drawn head or an image of a real person. Figures 13d and 13e illustrate how an embodiment with a hand-drawn character might be used in character animation. In particular, Figure 13d shows a hand-drawn sketch of a character which, when combined with the images from the source video sequence (some of which are shown in Figure 13a)

generate a target video sequence, some frames of which are shown in Figure 13e. As can be seen from a comparison of the corresponding frames in the source and target video frames, the hand-drawn sketch has been animated automatically using this technique. As those skilled in the art will appreciate, this is a much quicker and simpler technique for achieving computer animation as compared with existing systems which require the animator to manually create each frame of the animation. In particular, in this embodiment, all that is required is a video sequence of a real life actor acting out the scene to be animated, together with a single sketch of the character to be animated.

The above embodiment has described the way in which a target image can be used to modify a source video sequence. In order to do this, a set of appearance parameters has to be automatically calculated for each frame in the video sequence. This involved the use of a number of Active matrices which relate image errors to appearance parameter changes. As those skilled in the art will appreciate, similar processing is required in other applications, such as the tracking of an object within a video sequence, the tracking of a human face within a video sequence or the tracking of a knee joint in an MRI scan.

In the above embodiment, the appearance model was used to

model the variations in facial expressions and 3D pose of human heads. As those skilled in the art will appreciate, the appearance model can be used to model the appearance of any deformable object such as parts of the body and other animals and objects. For example, the above techniques can be used to track the movement of lips in a video sequence. Such an embodiment could be used in film dubbing applications in order to synchronise the lip movements with the dubbed sound. This animation technique might also be used to give animals and other objects human-like characteristics by combining images of them with a video sequence of an actor. This technique can also be used for monitoring the shape and appearance of objects passing along a production line for quality control purposes.

In the above embodiment, the appearance model was generated by using a principal component analysis of shape and texture data which is extracted from the training images. As those skilled in the art will appreciate, by modelling the features of the training heads in this way, it is possible to accurately model each head by just a small number of parameters. However, other modelling techniques, such as vector quantisation and wavelet techniques can be used.

In the above embodiments, the training images used to generate the appearance model were all colour images in

which each pixel had an RGB value. As those skilled in the art will appreciate, the way in which the colour is represented in this embodiment is not important. In particular, rather than each pixel having a red, green and blue value, they might be represented by a chrominance and a luminance component or by hue, saturation and value components. Alternatively still, the training images may be black and white images, in which case only grey level data would be extracted from the facets in the training images. Additionally, the resolution of each training image may be different.

In the above embodiment, during the automatic generation of the appearance parameters, and in particular during the iterative updating of these appearance parameters the error between the input image and the model image was generated using the appearance model. Since this iterative technique still requires a relatively accurate initial estimate for the appearance parameters, it is possible initially to perform the iterations using lower resolution images and once convergence has been reached for the lower resolutions to then increase the resolution of the images and to repeat the iterations for the higher resolutions. In such an embodiment, separate Active matrices would be required for each of the resolutions.

In the above embodiment, the difference parameters were determined by comparing the image of the first actor from

one of the frames of the source video sequence with the image of the second actor in the target image. In an alternative embodiment, a separate image of the first actor may be provided which does not form part of the source video sequence.

In the above embodiments, each of the appearance models modelled variations in two-dimensional images. The above modelling technique could be adapted to work with 3D images and animations. In such an embodiment, the training images used to generate the appearance model would normally include 3D images instead of 2D images. The three-dimensional models may be obtained using a three dimensional scanner which typically work either by using laser range-finding over the object or by using one or more stereo pairs of cameras. Once a 3D hierarchical appearance model has been created from the training models, new 3D models can be generated by adjusting the appearance parameters and existing 3D models can be animated using the same differencing technique that was used in the two-dimensional embodiment described above. This 3D model can then be used to track 3D objects directly within a 3D animation. Alternatively, a 2D model may be used to track the 3D object within a video sequence and then use the result to generate 3D data for the tracked object.

In the above embodiment, a set of difference parameters

were identified which describe the main differences between the head in the video sequence and the head in the target image, which difference parameters were used to modify the video sequence so as to generate a target video sequence showing the second head. In the embodiment, the set of difference parameters were added to a set of appearance parameters for the current frame being processed. In an alternative embodiment, the difference parameters may be weighted so that, for example, the target video sequence shows a head having characteristics from both the first and second actors.

In the above embodiment, a hierarchical appearance model is used to model the appearance of human faces. The model is then used to modify a source video sequence showing a first actor performing a scene to generate a target video sequence showing a second actor performing the same scene. As those skilled in the art will appreciate, the hierarchical model presented above can be used in various other applications. For example, the hierarchical appearance model can be used for synthetic two-dimensional or three-dimensional character generation; video compression when the video is substantially that of an object which is modelled by the appearance model; object recognition for security purposes; face tracking for human performance analysis or human computer interaction and the like; 3D model generation from two-dimensional images; and image editing

(for example making people look older or younger, fatter or thinner etc).

In the above embodiment, an iterative process was used to update an estimated set of appearance parameters for an input image. This iterative process continued until an error between the actual image and the image predicted by the model was below a predetermined threshold. In an alternative embodiment, where there is only a predetermined amount of time available for determining a set of appearance parameters for an input image, this iterative routine may be performed for a predetermined period of time or for a predetermined number of iterations.

CLAIMS:

1. A parametric model for modelling the shape of an object, the model comprising:

data defining a function which relates a set of input parameters to a set of locations which identify the relative positions of a plurality of predetermined points on the object;

characterised in that said data defines a hierarchical set of functions in which a function in a top layer of the hierarchy is operable to generate a set of output parameters from a set of input parameters and in which one or more functions in a bottom layer of the hierarchy are operable to receive parameters output from one or more functions from a higher layer of the hierarchy and to generate therefrom at least some of said locations which identify the relative positions of said predetermined points.

2. A model according to claim 1, wherein said hierarchy comprises one or more intermediate layers of functions which are operable to receive parameters output from one or more functions from a higher layer of the hierarchy and to generate therefrom a set of output parameters for input to functions in a lower layer of the hierarchy.

3. A model according to claim 1 or 2, for modelling the two-dimensional shape of the object by identifying the

relative positions of said predetermined points in a predetermined plane.

4. A model according to claim 1 or 2, for modelling the three-dimensional shape of the object by identifying the relative positions of the predetermined points in a three-dimensional space.

5. A model according to any preceding claim, wherein one or more of said functions comprises a linear function which linearly relates the input parameters to the function to the output parameters of the function.

6. A model according to claim 5, wherein said one or more linear functions are identified from a principal component analysis of training data derived from a set of training objects.

7. A model according to any preceding claim, wherein one or more of said functions are non-linear.

8. A model according to claim 7, wherein at least one of said non-linear functions comprises a neural network.

9. A model according to any preceding claim, wherein the number of parameters input to at least one of said functions is smaller than the number of parameters output from the function.

10. A model according to any preceding claim, wherein the number of input parameters to at least one of said functions is greater than or equal to the number of parameters output by the function.

11. A model according to any preceding claim for modelling the shape and texture of the object, the model further comprising data defining a hierarchical set of functions in which a function in a top layer of the hierarchy is operable to generate a set of output parameters from a set of input parameters and in which one or more functions in a bottom layer of the hierarchy are operable to receive parameters output from one or more functions from a higher layer of the hierarchy and to generate therefrom texture information for the object.

12. A model according to claim 11, wherein the texture hierarchy has the same structure as the shape hierarchy.

13. A model according to claim 11 or 12, wherein one or more of said functions are operable to relate an input set of shape and texture parameters to an output set of appearance parameters defining both shape and texture.

14. A model according to any preceding claim, wherein said object is a deformable object.

15. A model according to claim 14, wherein said

deformable object includes a human face.

16. A model according to claim 15, wherein said function in said top layer of the hierarchy models the shape of the entire face and wherein said hierarchy includes a function which models the shape of the mouth.

17. A model according to claim 16, wherein said hierarchy further comprises a function for modelling the shape of the eyes.

18. A model according to any preceding claim, wherein the or each function in the bottom layer of the hierarchy identifies the positions of a plurality of predetermined points according to a predefined function of smaller number of control point positions.

19. A model according to claim 18, wherein the predefined function for each of the plurality of points is a linear mapping of the control point positions and the control points are the three corners of a triangular facet.

20. A model according to claim 18, wherein the predefined function for each of the plurality of points is a predefined non-linear mapping of a fixed number of control point positions.

21. A model according to claim 18, wherein the predefined function for each of the plurality of points is a predefined displacement from a single control point.

22. A method of determining a set of appearance parameters representative of the appearance of an object, the method comprising the steps of:

(i) storing a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object;

(ii) storing at least one function which relates a change in the input parameters to an error between actual appearance data for the object and appearance data determined from the set of input parameters and said parametric model;

(iii) initially estimating a current set of input parameters for the object;

(iv) determining appearance data for the object from the current set of input parameters and the stored parametric model;

(v) determining the error between actual appearance data of the object and the appearance data determined from the current set of input parameters;

(vi) determining a change in the input parameters using said at least one stored function and said determined error; and

(vii) updating the current set of input parameters

with the determined change in the input parameters.

23. A method according to claim 22, further comprising the step of repeating steps (iv) to (vii) until the error determined in step (v) is less than a predetermined threshold.

24. A method according to claim 22, further comprising the step of repeating steps (iv) to (vii) for a predetermined amount of time or for a predetermined number of repetitions.

25. A method according to claim 22, 23 or 24, wherein said second storing step stores a plurality of functions, one associated with each function within the hierarchical model.

26. A method of tracking an object comprising the steps of:

(i) storing a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object;

(ii) storing at least one function which relates a change in the input parameters to an error between the actual appearance data for the object and the appearance data determined from the set of input parameters and said parametric model;

(iii) initially estimating a current set of input parameters for the object;

(iv) determining the appearance data for the object from the current set of input parameters and the stored parametric model;

(v) determining an error between the actual appearance data for the object and the appearance data for the object determined from the current set of input parameters;

(vi) determining a change in the input parameters using the at least one stored function and the determined error;

(vii) updating the current set of input parameters with said change in the input parameters;

(viii) repeating steps (iv) to (vii) in order to reduce the error determined in step (v); and

(ix) repeating steps (iii) to (viii) to track the object.

27. An apparatus for determining a set of appearance parameters representative of the appearance of an object, the apparatus comprising:

means for storing (i) a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object; and (ii) at least one function which relates a change in the input parameters to an error between actual appearance data for the object and

the appearance data for the object determined from the set of input parameters and said parametric model;

means for receiving an initial estimate of a current set of input parameters for the object;

means for updating the current set of input parameters comprising:

(i) means for determining appearance data for the object from the current set of input parameters and the stored parametric model;

(ii) means for determining the error between the actual appearance data for the object and the appearance data for the object determined from the current set of input parameters;

(iii) means for determining a change in the input parameters using said at least one stored function and said determined error; and

(iv) means for updating the current set of input parameters with the determined change in the input parameters.

28. An apparatus according to claim 27, wherein said updating means is operable to update iteratively the current set of input parameters until the error determining means determines an error which is less than a predetermined threshold.

29. An apparatus according to claim 27 or 28, wherein said storing means stores a plurality of functions, one

associated with each function within the hierarchical model.

30. An apparatus for tracking an object comprising:

means for storing (i) a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object; and (ii) at least one function which relates a change in the input parameters to an error between actual appearance data for the object and the appearance data for the object determined from the set of input parameters and said parametric model;

means for receiving an initial estimate of a current set of input parameters for the object;

means for updating the current set of input parameters comprising:

(i) means for determining appearance data for the object from the current set of input parameters and the stored parametric model;

(ii) means for determining an error between actual appearance data for the object and the appearance data for the object determined from the current set of input parameters;

(iii) means for determining a change in the input parameters using the at least one stored function and the determined error; and

(iv) means for updating the current set of input parameters with said change in the input parameters;

wherein said updating means is operable to update iteratively the current set of input parameters in order to reduce the determined error, wherein said receiving means is operable to receive further estimates of the current input parameters and wherein said update means is operable to update the received estimates of the current input parameters in order to track said object.

31. A storage medium storing the parametric model according to any of claims 1 to 21 or storing processor implementable instructions for controlling a processor to implement the method of any one of claims 22 to 26.

32. Processor implementable instructions for controlling a processor to implement the method of any one of claims 22 to 26.

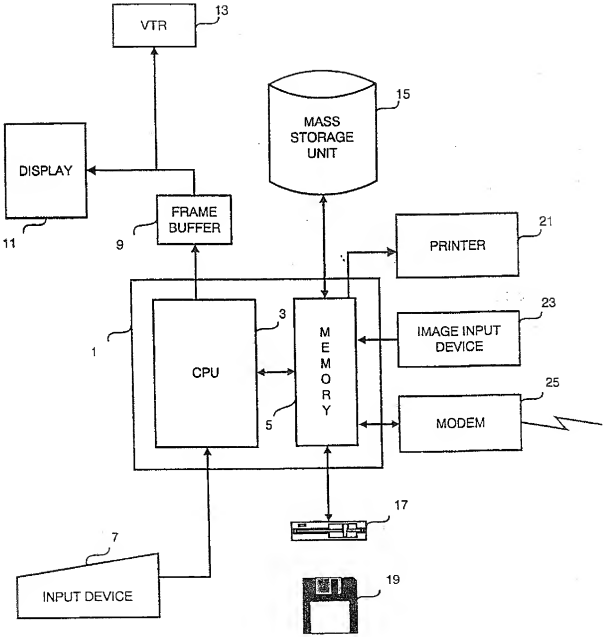

VTR
13
MASS STORAGE UNIT
15
DISPLAY
11
FRAME BUFFER
9
PRINTER
21
1
CPU
3
MEMORY
5
IMAGE INPUT DEVICE
23
MODEM
25
17
7
INPUT DEVICE
19

Fig. 1

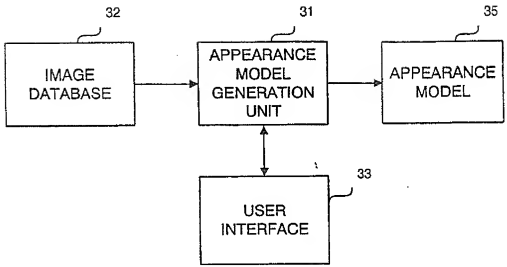

32
31
35
IMAGE DATABASE
APPEARANCE MODEL GENERATION UNIT
APPEARANCE MODEL
33
USER INTERFACE

Fig. 2

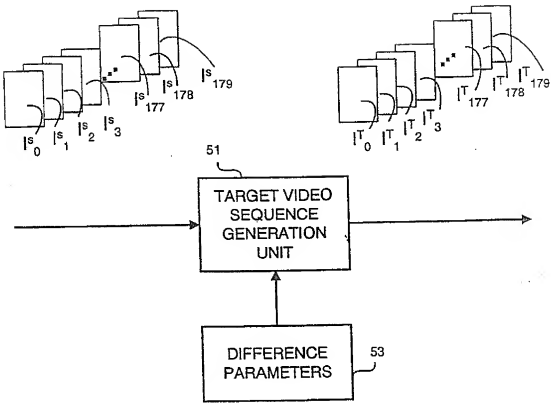

I^S_0 I^S_1 I^S_2 I^S_3 I^S_{177} I^S_{178} I^S_{179}
I^T_0 I^T_1 I^T_2 I^T_3 I^T_{177} I^T_{178} I^T_{179}
51
TARGET VIDEO SEQUENCE GENERATION UNIT
DIFFERENCE PARAMETERS
53

Fig. 3

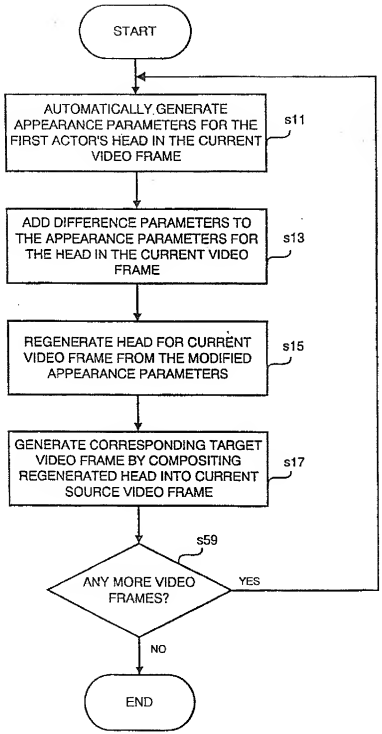

START
AUTOMATICALLY GENERATE APPEARANCE PARAMETERS FOR THE FIRST ACTOR'S HEAD IN THE CURRENT VIDEO FRAME
s11
ADD DIFFERENCE PARAMETERS TO THE APPEARANCE PARAMETERS FOR THE HEAD IN THE CURRENT VIDEO FRAME
s13
REGENERATE HEAD FOR CURRENT VIDEO FRAME FROM THE MODIFIED APPEARANCE PARAMETERS
s15
GENERATE CORRESPONDING TARGET VIDEO FRAME BY COMPOSITING REGENERATED HEAD INTO CURRENT SOURCE VIDEO FRAME
s17
s59
ANY MORE VIDEO FRAMES?
YES
NO
END

Fig. 4

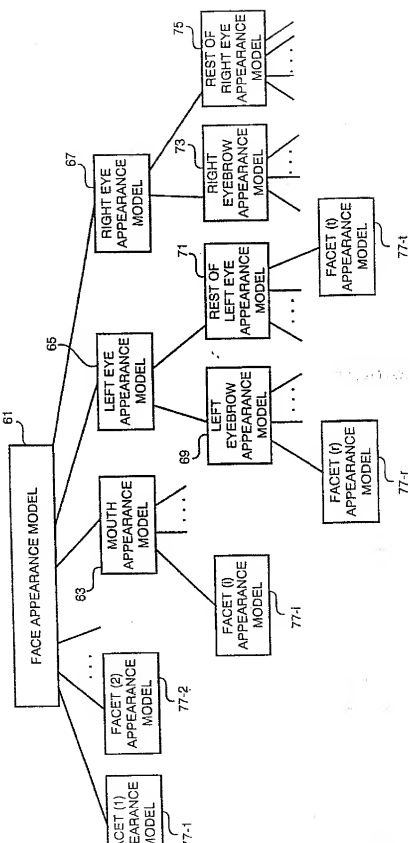

FACE APPEARANCE MODEL
61
FACET (1) APPEARANCE MODEL
77-1
FACET (2) APPEARANCE MODEL
77-2
FACET (i) APPEARANCE MODEL
77-i
MOUTH APPEARANCE MODEL
63
LEFT EYE APPEARANCE MODEL
65
LEFT EYEBROW APPEARANCE MODEL
69
FACET (r) APPEARANCE MODEL
77-r
REST OF LEFT EYE APPEARANCE MODEL
71
FACET (t) APPEARANCE MODEL
77-t
RIGHT EYE APPEARANCE MODEL
67
RIGHT EYEBROW APPEARANCE MODEL
73
REST OF RIGHT EYE APPEARANCE MODEL
75

Fig. 5

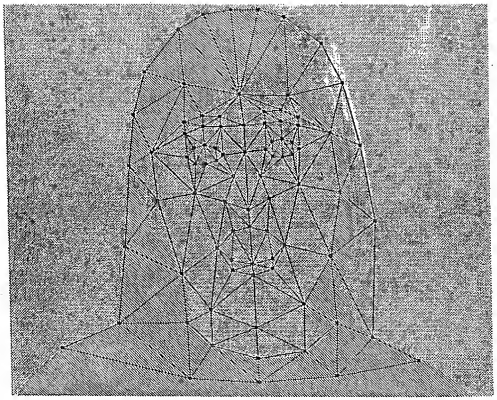


Fig. 6

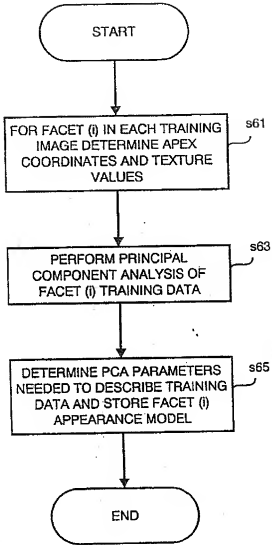

START
FOR FACET (i) IN EACH TRAINING IMAGE DETERMINE APEX COORDINATES AND TEXTURE VALUES
s61
PERFORM PRINCIPAL COMPONENT ANALYSIS OF FACET (i) TRAINING DATA
s63
DETERMINE PCA PARAMETERS NEEDED TO DESCRIBE TRAINING DATA AND STORE FACET (i) APPEARANCE MODEL
s65
END

Fig. 7

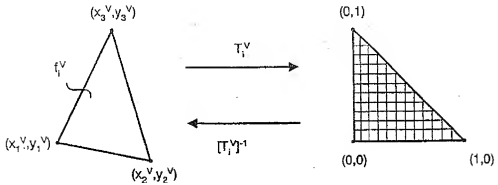


Fig. 8

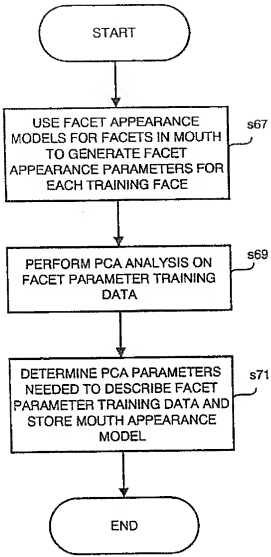

START
USE FACET APPEARANCE MODELS FOR FACETS IN MOUTH TO GENERATE FACET APPEARANCE PARAMETERS FOR EACH TRAINING FACE
s67
PERFORM PCA ANALYSIS ON FACET PARAMETER TRAINING DATA
s69
DETERMINE PCA PARAMETERS NEEDED TO DESCRIBE FACET PARAMETER TRAINING DATA AND STORE MOUTH APPEARANCE MODEL
s71
END

Fig. 9

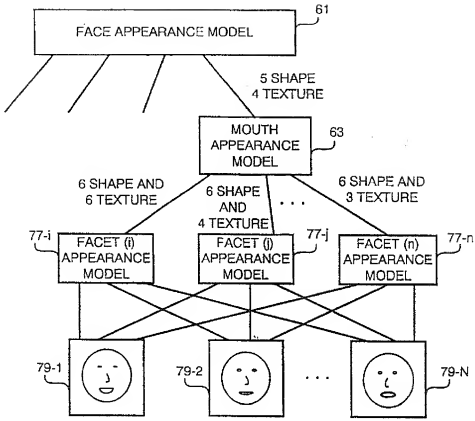

FACE APPEARANCE MODEL
61
5 SHAPE
4 TEXTURE
MOUTH
APPEARANCE
MODEL
63
6 SHAPE AND
6 TEXTURE
6 SHAPE
AND
4 TEXTURE
6 SHAPE AND
3 TEXTURE
77-i
FACET (i)
APPEARANCE
MODEL
77-j
FACET (j)
APPEARANCE
MODEL
FACET (n)
APPEARANCE
MODEL
77-n
79-1
79-2
79-N

Fig. 10

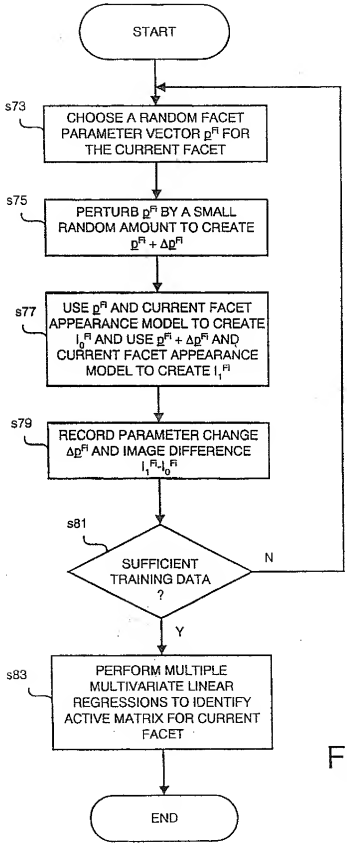

START
s73
CHOOSE A RANDOM FACET PARAMETER VECTOR $\underline{p}^{Fi}$ FOR THE CURRENT FACET
s75
PERTURB $\underline{p}^{Fi}$ BY A SMALL RANDOM AMOUNT TO CREATE $\underline{p}^{Fi} + \Delta\underline{p}^{Fi}$
s77
USE $\underline{p}^{Fi}$ AND CURRENT FACET APPEARANCE MODEL TO CREATE I_0^{Fi} AND USE $\underline{p}^{Fi} + \Delta\underline{p}^{Fi}$ AND CURRENT FACET APPEARANCE MODEL TO CREATE I_1^{Fi}
s79
RECORD PARAMETER CHANGE $\Delta\underline{p}^{Fi}$ AND IMAGE DIFFERENCE $I_1^{Fi} - I_0^{Fi}$
s81
SUFFICIENT TRAINING DATA ?
N
Y
s83
PERFORM MULTIPLE MULTIVARIATE LINEAR REGRESSIONS TO IDENTIFY ACTIVE MATRIX FOR CURRENT FACET
END

Fig. 11a

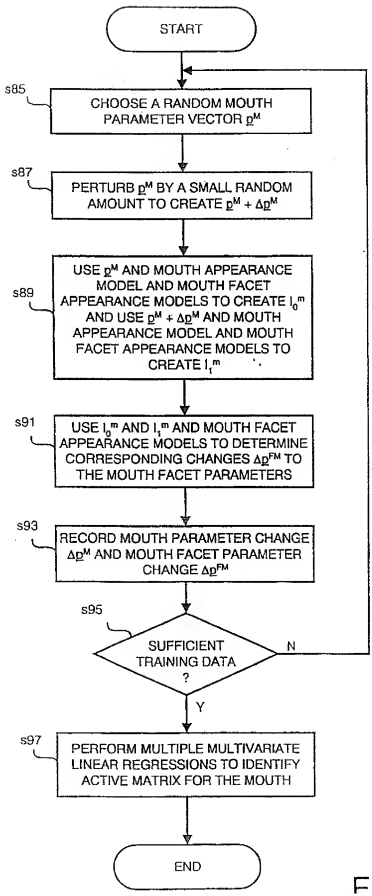

START
s85
CHOOSE A RANDOM MOUTH PARAMETER VECTOR $\underline{p}^M$
s87
PERTURB $\underline{p}^M$ BY A SMALL RANDOM AMOUNT TO CREATE $\underline{p}^M + \Delta\underline{p}^M$
s89
USE $\underline{p}^M$ AND MOUTH APPEARANCE MODEL AND MOUTH FACET APPEARANCE MODELS TO CREATE I_0^m AND USE $\underline{p}^M + \Delta\underline{p}^M$ AND MOUTH APPEARANCE MODEL AND MOUTH FACET APPEARANCE MODELS TO CREATE I_1^m
s91
USE I_0^m AND I_1^m AND MOUTH FACET APPEARANCE MODELS TO DETERMINE CORRESPONDING CHANGES $\Delta\underline{p}^{FM}$ TO THE MOUTH FACET PARAMETERS
s93
RECORD MOUTH PARAMETER CHANGE $\Delta\underline{p}^M$ AND MOUTH FACET PARAMETER CHANGE $\Delta\underline{p}^{FM}$
s95
SUFFICIENT TRAINING DATA ?
N
Y
s97
PERFORM MULTIPLE MULTIVARIATE LINEAR REGRESSIONS TO IDENTIFY ACTIVE MATRIX FOR THE MOUTH
END

Fig. 11b

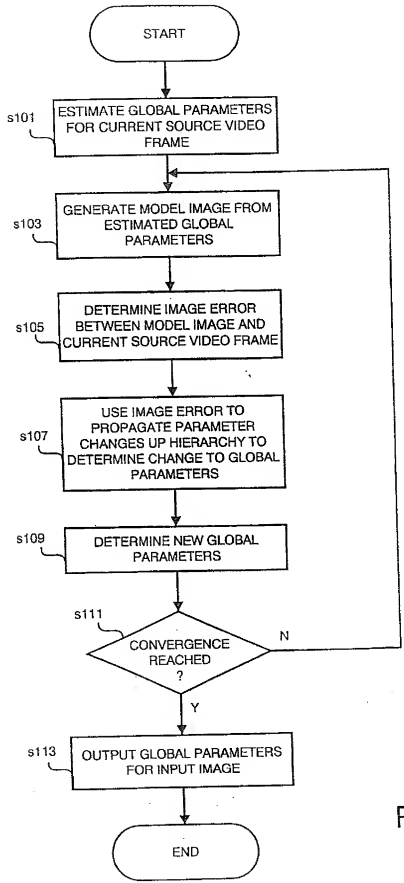

START
s101
ESTIMATE GLOBAL PARAMETERS FOR CURRENT SOURCE VIDEO FRAME
s103
GENERATE MODEL IMAGE FROM ESTIMATED GLOBAL PARAMETERS
s105
DETERMINE IMAGE ERROR BETWEEN MODEL IMAGE AND CURRENT SOURCE VIDEO FRAME
s107
USE IMAGE ERROR TO PROPAGATE PARAMETER CHANGES UP HIERARCHY TO DETERMINE CHANGE TO GLOBAL PARAMETERS
s109
DETERMINE NEW GLOBAL PARAMETERS
s111
CONVERGENCE REACHED ?
N
Y
s113
OUTPUT GLOBAL PARAMETERS FOR INPUT IMAGE
END

Fig. 12

Fig. 13a

Fig. 13b

Fig. 13c

Fig. 13d

INTERNATIONAL SEARCH REPORT

Internal Application No
PCT/GB 00/04411

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06T17/00 G06K9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06K G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, INSPEC, IBM-TDB, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EDWARDS G J ET AL: "ADVANCES IN ACTIVE APPEARANCE MODELS" KERKYRA, GREECE, SEPT. 20 - 27, 1999,LOS ALMITOS, CA: IEEE COMP. PRESS,US, vol. CONF. 7, 1999, pages 137-142, XP000980072 ISBN: 0-7695-0165-6 abstract page 137, left-hand column, paragraph 1 -right-hand column, paragraph 2 --- -/--	1-32

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

A document defining the general state of the art which is not considered to be of particular relevance

E earlier document but published on or after the international filing date

L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

O document referring to an oral disclosure, use, exhibition or other means

P document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 April 2001	04/05/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	König, W

INTERNATIONAL SEARCH REPORT

Interna .al Application No
PCT/GB 00/04411

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	COOTES T F ET AL: "ACTIVE SHAPE MODELS-THEIR TRAINING AND APPLICATION" COMPUTER VISION AND IMAGE UNDERSTANDING, ACADEMIC PRESS, US, vol. 61, no. 1, January 1995 (1995-01), pages 38-59, XP000978654 ISSN: 1077-3142 abstract page 38, left-hand column, paragraph 1 -page 39, left-hand column, paragraph 1	1-32